

Zur Anwendung der Interaction-Coding-Technik

Prüfer, Peter; Rexroth, Margrit

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Prüfer, P., & Rexroth, M. (1985). Zur Anwendung der Interaction-Coding-Technik. *ZUMA Nachrichten*, 9(17), 2-49.
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210344>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Zur Anwendung der Interaction-Coding-Technik

1. Einleitende Bemerkungen

In der sozialwissenschaftlichen Umfrageforschung gewinnt die Diskussion um die sog. "Qualität der Daten" immer mehr an Gewicht. Gegenstand dieser Diskussion ist die in den letzten Jahren verstärkt erhobene Forderung nach validen Umfragedaten. Was den methodischen Aspekt dieser Forderung betrifft, so rückt damit der Datenerhebungsprozeß in den Mittelpunkt des Interesses, da besonders hier Effekte wirksam werden können, die sich negativ auf die Qualität der Daten auswirken können. Diese Effekte können zum einen durch den Interviewer verursacht werden, der z.B. durch unangemessenes Verhalten Reaktionen beim Befragten hervorruft, die der eigentlichen Fragenintention nicht gerecht werden, zum anderen können auch unklar formulierte Fragen die Datenqualität negativ beeinflussen.

Bisher hat man versucht, den durch Interviewer und Instrument verursachten Effekten durch entsprechende Interviewer-Schulungen und Pretests entgegenzuwirken mit dem Ziel, evtl. vorhandene Mängel am Instrument aufzudecken und eine möglichst optimale Ausgangssituation für die Erhebung der Daten zu schaffen. In diesem Zusammenhang schien die Erprobung einer Technik interessant, die durch die Beobachtung des gesamten Interviewverlaufs Aussagen sowohl über Interviewer als auch über das Instrument zuläßt. Dabei handelt es sich um die sog. Interaction-Coding-Technik, deren Grundlagen auf Charles CANNELL et al. (1975) zurückgehen.

Der vorliegende Beitrag befaßt sich mit der Frage, in welchem Umfang diese Technik in der Lage ist, einen Beitrag zur Verbesserung der Qualität von Umfragedaten zu liefern. Dazu wurden zwei separate Untersuchungen durchgeführt: In der ersten Untersuchung sollte die Frage geklärt werden, inwieweit die Interaction-Coding-Technik geeignet ist, Interviewerverhalten zu registrieren, zu analysieren und zu bewerten. In der zweiten Untersuchung steht das Instrument, der Fragebogen bzw. stehen die einzelnen Fragen im Mittelpunkt. Geprüft werden sollte hier, ob neben herkömmlichen Pretestmethoden die Interaction-Coding-Technik zusätzliche Befunde zum "Funktionieren" einzelner Fragen liefert, die über diejenigen konventioneller Techniken hinausgehen.

ZUMA

Was den Interviewer betrifft, so kann zunächst festgestellt werden, daß dieser bereits Gegenstand unzähliger Forschungsarbeiten war. Dies ist ein Ausdruck dafür, welche wichtige Rolle der Interviewer im Forschungsprozeß der empirischen Sozialforschung einnimmt. Soll er doch als sog. "Agent" des Forschers mit Hilfe des Instruments "Fragebogen" Daten in Form valider Antworten der Befragten sammeln. Nicht ohne Grund also wird der sog. "Survey-Interviewer" von Charles CANNELL (1981) als "gatekeeper" zu den Einstellungen, Erfahrungen und Wahrnehmungen der Befragten bezeichnet.

Die in der Literatur erwähnten Forschungsarbeiten zeigen übereinstimmend, daß die Antworten der Befragten auf vielfältige Weise von der Person des Interviewers beeinflusst werden können. Bis hin zu RICE (1929) läßt sich die lange Tradition der Untersuchungen zu diesen sog. "Interviewereffekten" zurückverfolgen. Die Ergebnisse machen deutlich, daß diese Effekte, d.h. die unerwünschte Veränderung der Daten, einmal durch äußere Merkmale des Interviewers, wie z.B. Geschlecht, Alter oder Hautfarbe, zum anderen durch sein Verhalten hervorgerufen werden können. So konnte z.B. nachgewiesen werden, daß die Einstellungen des Interviewers zu den im Fragebogen angesprochenen Themen durchaus Einfluß auf die Antworten der Befragten haben können. Die Beeinflussung des Befragten kann in diesem Fall z.B. hervorgerufen werden durch vom Interviewer benutzte sog. "leading probes", d.h. Nachfragen, die den Befragten in eine ganz bestimmte Richtung lenken, durch Unterlassen von Nachfragen, durch fehlerhaftes Notieren der Antworten oder durch andere nonverbale Verhaltensweisen.

Die Qualität der Daten wäre in einem solchen Fall erheblich gemindert, verursacht durch "falsches" Interviewerverhalten. Vermeiden bzw. reduzieren läßt sich diese Art von Interviewereffekten durch geeignete Schulungsmaßnahmen. Dazu sind zwei Schritte notwendig:

1. Die Erfassung von "falschem" Interviewerverhalten, d.h. Verhalten, das von den vorgeschriebenen Regeln abweicht;
2. Interventionsmaßnahmen zur Beseitigung dieser "falschen" Verhaltensweisen.

Zur Realisierung dieser beiden Schritte scheint die Interaction-Coding-Technik geeignet. Die Frage, inwieweit diese Technik tatsächlich in der

Lage ist, die genannten Aufgaben zu lösen, war Gegenstand der ersten Untersuchung.

Wie bereits eingangs erwähnt, können neben dem Interviewerverhalten besonders das Instrument bzw. die einzelnen Fragen Ausgangspunkt für Effekte sein, die sich negativ auf die Datenqualität auswirken. Die Feststellung der Güte einer Frage ist jedoch nicht unproblematisch, da kaum objektive, empirisch überprüfte Kriterien hierzu existieren. Diese unbefriedigende Situation wird von CANNELL et al. (1985) folgendermaßen charakterisiert:

"The least scientifically rigorous aspect of survey research is the development and testing of questions. It is ironic that the creation of the measuring instrument is based primarily on past experience with only a few 'common sense' principles as guidance".

In der Umfrageforschung ist es bislang üblich, Hinweise zur Qualität von Fragen aus einem der eigentlichen Befragung vorangehenden sog. Pretest zu erlangen. Dies geschieht in der Regel, indem erfahrene Interviewer eine Reihe von Interviews durchführen und mittels eines Berichts die bei einzelnen Fragen aufgetretenen Probleme an den Forscher weitervermitteln. Diese Mitteilungen des Interviewers sind sowohl subjektiv als auch unsystematisch und beschränken sich meistens nur auf schwerwiegende Probleme, die sich in der Interviewsituation ergeben. Die Grundlagen zur Konstruktion des endgültigen Instruments (= Fragebogen) bestehen daher überwiegend aus der subjektiven Beurteilung des Forschers und nur zu einem geringen Teil aus empirischen Befunden.

Es gibt in der Literatur Versuche, aus der eigenen Erfahrung abgeleitete allgemein gültige Regeln zur Konstruktion von Fragen aufzustellen. Oftmals bleibt es jedoch in solchen Fällen bei eher trivialen Hinweisen (z.B.: "Vermeiden Sie Fragen, die den Befragten in einer bestimmten Richtung beeinflussen können"). Ein frühes Standardwerk mit Regeln zur Fragenformulierung (PAYNE, 1951: "The art of asking questions") läßt schon vom Titel her Rückschlüsse auf die Art der im Buch beschriebenen insgesamt 100 Regeln zu. Bezeichnenderweise lautet der letzte Absatz:

"Actually you won't need this check list type of stimulus for long because most of these things are only common sense anyway. Having once been pointed

out, they should stay with you pretty well with perhaps only an occasional reading for a refresher".

Die in den letzten Jahren veröffentlichten systematischen Untersuchungen zu dieser Thematik, z.B. von SCHUMAN & PRESSER (1981) oder SUDMAN & BRADBURN (1982), die sich z.T. jedoch nur mit speziellen Fragetypen beschäftigen, zeigen deutliche Bestrebungen, die Konstruktion von Fragen auf eine empirische Grundlage zu stellen. Mit der hier vorgestellten zweiten Untersuchung soll ein Versuch unternommen werden, mit Hilfe der Interaction-Coding-Technik durch empirisch fundierte Aussagen auch einen Beitrag zur Bewertung der Qualität von Fragen zu liefern, da es Hinweise gab, daß aus der Interaktion zwischen Interviewer und Befragtem Rückschlüsse auf die Qualität von Fragen gezogen werden können.

2. Untersuchung 1 - Bewertung von Interviewerverhalten mit Hilfe der Interaction-Coding-Technik

2.1. Allgemeine Bemerkungen zum Bewertungssystem

Die erste Arbeit mit dem Interaction-Coding-System orientierte sich stark an dem bei CANNELL et al. (1975) beschriebenen System. Diese Technik ist so konzipiert, daß das Interviewerverhalten beim face-to-face-Interview analysiert und gleichzeitig bewertet werden kann. In ihrer Anwendung ist sie einfacher als die von den Autoren für spätere Arbeiten modifizierte und zur Bestimmung der Qualität von Fragen eingesetzte; somit bietet die bei CANNELL beschriebene Technik zur Bewertung von Interviewerverhalten einen guten Einstieg in ihre grundsätzliche Erprobung. Mit der ihr zugrundeliegenden Konzeption wird nicht der gesamte soziale Interaktionsprozeß zwischen Befragtem und Interviewer analysiert, sondern lediglich der Verhaltensbereich des Interviewers während der Erhebung. Die Technik konzentriert sich also in Untersuchung 1 nur auf einen der beiden Akteure.

2.2. Aufbau und Anwendung der Technik - das Codesystem

2.2.1. Aufbau der Technik

Die Voraussetzung zur Anwendung der Technik ist die Aufzeichnung der Interviews auf Tonträger. Zusätzlich werden vom Interviewer die Antworten des Befragten im Fragebogen registriert, so daß durch die Tonbandaufzeichnungen eine Kontrolle der Mitschriften möglich ist. Beim Abhören der Aufzeichnungen werden alle verbalen Aktivitäten des Interviewers während der Erhebung nach einem detaillierten Codesystem bewertet. Das Codesystem besteht

ZUMA

aus einer Aufzählung aller Verhaltensweisen (aller verbalen Aktivitäten), die ein Interviewer während des Interviews zeigen kann. Diese Verhaltensweisen sind in vier Bereiche gruppiert:

- I. Verhaltensbereich: beinhaltet alle Verhaltensweisen, die sich mit der Fragestellung beschäftigen;
- II. Verhaltensbereich: diejenigen Verhaltensweisen, die sich mit Klärung und Nachfrage befassen;
- III. Verhaltensbereich: umfaßt alle Verhaltensweisen, die als sonstige Verhaltensweisen zusammengefaßt werden;
- IV. Verhaltensbereich: beschäftigt sich mit der Einhaltung der vorgeschriebenen Fragenreihenfolge.

Innerhalb jedes dieser vier Verhaltensbereiche wird zwischen angemessenen und unangemessenen Verhaltensweisen unterschieden. Dieser Einteilung liegen bestimmte Interviewerverhaltensregeln zugrunde, die von der grundsätzlichen Konzeption des jeweiligen Fragebogentyps abgeleitet sind. Dabei ist zu beachten, daß auch beim standardisierten Interview nicht für alle möglichen Verhaltensweisen ein absolut gültiges Regelwerk festgeschrieben ist. Es gibt z.B. Verhaltensregeln zum Einstieg in ein Interview, zur Schaffung von Interviewatmosphäre, zum Bestätigungsverhalten dem Befragten gegenüber, die nach Auffassung des jeweiligen Forschers gehandhabt werden sollen. Demgegenüber stehen grundsätzlich festgeschriebene Verhaltensregeln, die als allgemein gültig angesehen werden können und dem Interviewer die Orientierung für seine Verhaltensweisen beim standardisierten Interview geben. Diese sind z.B.: Fragetext wörtlich vorlesen, Fragensukzession einhalten, Interviewer-Anweisungen strikt befolgen, absolute Neutralität usw.

Für einen Fragebogentyp mit weniger standardisiertem Charakter müßten dem Codesystem entsprechend variierte Verhaltensregeln zugrunde gelegt werden. Die hier vorgestellte Technik ist ausschließlich für das standardisierte Interview konzipiert. Aus pragmatischen Gründen ist jeder Verhaltensweise eine Codeziffer zugeordnet. Wie Tabelle 1 zeigt, unterliegt diese Zuordnung einer bestimmten Systematik.

ZUMA

Tabelle 1: Übersicht über die Codegruppen im Codeschema

	<u>Obergruppe =</u> <u>Codegruppe</u>	<u>Untergruppe =</u> <u>Codeziffern</u>
I. FRAGESTELLUNG		
angemessene Verhaltensweisen (korrekte Fragestellung)	10	11-13
unangemessene Verhaltensweisen (nicht korrekte Fragestellung)	20	21-23
II. KLÄRUNG/NACHFRAGE		
angemessene Verhaltensweisen (Klärung oder Nachfrage auf nicht direkte Weise)	30	31-36
unangemessene Verhaltensweisen (Klärung oder Nachfrage auf direktive Weise)	40	41-47
III. SONSTIGES VERHALTEN		
sonstiges angemessenes Verhalten	50	51,58
sonstiges unangemessenes Verhalten (verbal)	60	62-68
sonstiges unangemessenes Verhalten (nicht-verbal)	70	71-75
IV. FILTER		
korrekter Filtersprung	80	81
Filterfehler	90	91,92

So werden z.B. für den I. Bereich "Fragestellung" alle angemessenen Verhaltensweisen in der Codegruppe 10 zusammengefaßt. Innerhalb dieser 10er Gruppe symbolisieren die Codes 11, 12 und 13 angemessene korrekte Verhaltensweisen beim Vorlesen des Fragetextes. Alle unangemessenen Verhaltensweisen dieses Verhaltensbereichs werden innerhalb der Codegruppe 20 zusammengefaßt. Hier symbolisieren die Codes 21, 22, 23 unangemessene Verhaltensweisen beim Vorlesen des Fragetextes.

2.2.2. Vergabe von Codeziffern

Jeder der auf der Bändaufzeichnung registrierten Aktivitäten des Interviewers wird - unabhängig davon, ob angemessen oder unangemessen - eine entsprechende Codeziffer zugeordnet. Dabei werden die Codeziffern fragenspezifisch notiert, wobei die Anzahl der pro Frage vergebenen Codes vom Interaktionsprozeß zwischen Interviewer und Befragtem abhängt. Bei der Durchführung des standardisierten Interviews konzentriert sich die Hauptaufgabe des Interviewers auf das wörtliche Vorlesen des Fragetextes und das korrekte Registrieren der Antwort. Im Idealfall wird also pro Frage nur eine Codeziffer vergeben, die das korrekte wörtliche Vorgeben des Fragetextes bewertet (Code 11). Eine Ausnahme bildet der Filterhinweis, der durch einen weiteren Code aus der 80er oder 90er Gruppe belegt ist. Treten jedoch bei Fragenunverständnis Rückfragen durch den Befragten auf oder ordnet sich dieser nicht eindeutig den Antwortvorgaben zu, so sind Zusatzerkklärungen durch den Interviewer nötig, die jeweils durch entsprechende Codes bewertet werden. Je ausführlicher die Kommunikation zwischen Interviewer und Befragtem, desto mehr Codes kommen zur Auswertung. Die in diesem Codesystem definierten Verhaltensregeln beziehen sich primär auf akustisch registrierte Aktivitäten. Eine Ausnahme bildet die Codegruppe 70, die unangemessenes nicht-verbales Verhalten erfaßt. Konkret handelt es sich dabei um Verhaltensweisen wie z.B.:

- Der Interviewer notiert die bereits während des Interviewverlaufs erhaltene Antwort, ohne die Frage noch einmal zu stellen;
- der Interviewer unterläßt bei unzulänglicher Antwort eine klärende Nachfrage;
- der Interviewer versäumt, bei falschem Verständnis den Sinn der Frage richtigzustellen.

Im folgenden soll die Anwendung und inhaltliche Bedeutung einzelner Codes an Beispielen verdeutlicht werden; im Anschluß daran soll anhand einer "live"-Erhebungssituation (Pretestinterview Wohlfahrtssurvey 84, Frage 33) demonstriert werden, wie durch unangemessenes Verhalten des Interviewers und bedingt durch die Reaktionen des Befragten die Vergabe einer Vielzahl von Codes für den Interviewer notwendig wird.

2.2.2.1. Beispiele für die Bedeutung einzelner Codes

CODE 12 (angemessenes Verhalten im Bereich Fragestellung, Codegruppe 10)

Bedeutung:

Interviewer liest Fragetext geringfügig verändert vor, ohne den Bezugsrahmen zu verändern; es werden keine Schlüsselworte hinzugefügt, weggelassen oder verändert.

Fragetext:

1. Sagen Sie mir bitten anhand dieser Liste, von wem dieser Kurs veranstaltet wurde.

2. Welchem der folgenden Bereiche auf dieser Liste würden Sie diesen Lehrgang oder Kurs zuordnen?

Interviewer:

Sagen Sie mir bitte anhand dieser Liste /hier/, von wem dieser Kurs veranstaltet wurde.

/Die nächste Frage lautet: /
Welchem der folgenden Bereiche auf dieser Liste würden Sie diesen Lehrgang oder Kurs zuordnen?

Erläuterung:

Der Interviewer nimmt geringfügige Textveränderungen vor, die keine inhaltlichen Konsequenzen haben und somit zulässig sind.

CODE 22 (unangemessenes Verhalten im Bereich Fragestellung, Codegruppe 20)

Bedeutung:

Interviewer verändert den Fragetext erheblich; es werden Schlüsselworte hinzugefügt, weggelassen oder verändert.

Fragetext:

1. Wie hoch ist das Netto-Einkommen, das Sie in diesem Haushalt hier alle zusammen haben?

Interviewer:

Wie hoch ist /ungefähr/ das Netto-Einkommen, das Sie in diesem Haushalt hier alle zusammen haben?

Erläuterung:

Durch Hinzufügung des Wortes "ungefähr" wird die eigentliche Zielrichtung der Frage verändert. Der Interviewer versucht hier, die Peinlichkeit der Situation zu mildern, indem er dem Befragten einen gewissen Beantwortungsspielraum zugesteht. Diese Beobachtung ist häufig bei den als "heikel" geltenden Fragen zu machen.

CODE 34 (angemessenes Verhalten im Bereich Klärung/Nachfrage,
Codegruppe 30)

Bedeutung:

Interviewer wiederholt oder klärt Antwort des Befragten korrekt auf nicht direkte Weise.

Fragetext:

... Finden Sie diese Möglichkeit für sich selbst sehr gut, gut, weniger gut oder überhaupt nicht gut?

Interviewer:

... Finden Sie diese Möglichkeit für sich selbst sehr gut, gut, weniger gut oder überhaupt nicht gut?

Befragter:

Oh, finde ich ganz gut.

Interviewer:

Finden Sie das sehr gut oder gut?

Erläuterung:

Der Interviewer zeigt hier insofern ein korrektes Verhalten, als er zur Präzisierung der Antwort des Befragten die im Fragetext vorgegebenen Antwortkategorien noch einmal vorgibt und somit den Befragten auffordert, sich für eine Vorgabe präzise zu entscheiden. Es werden Code 11 für das korrekte Vorlesen des Fragetextes (= 1. Aktivität des Interviewers) und Code 34 für Klärung der Antwort (= 2. Aktivität des Interviewers) vergeben.

CODE 44 (unangemessenes Verhalten im Bereich Klärung/Nachfrage,
Codegruppe 40)

Bedeutung:

Interviewer gibt eine von der tatsächlichen Antwort des Befragten abweichende Zusammenfassung oder ordnet die Antwort eigenmächtig den vorgegebenen Antwortkategorien zu.

Fragetext:

Wenn Sie nun an Ihre eigene Situation denken, was halten Sie dann von der Möglichkeit, ein Studium neben dem Beruf wahrzunehmen? Finden Sie diese Möglichkeit für sich selbst sehr gut, gut, weniger gut oder überhaupt nicht gut?

Interviewer:

Wenn Sie nun an Ihre eigene Situation denken, was halten Sie dann von der Möglichkeit, ein Studium neben dem Beruf wahrzunehmen? Finden Sie diese Möglichkeit für sich selbst sehr gut, gut, weniger gut oder überhaupt nicht gut?

Befragter:

Das ist etwas für junge Leute, ich glaube, ich bin dazu schon zu alt.

Interviewer:

Also, Sie finden das weniger gut.

Erläuterung:

Der Interviewer interpretiert die in eigenen Worten gegebene Antwort des Befragten und ordnet diese eigenmächtig den Antwortkategorien zu, ein direktives Interviewerverhalten, das manipulierte Daten produziert. Es werden Code 11 (1. Aktivität = korrektes Vorlesen des Fragetextes) und Code 44 (2. Aktivität = eigenmächtige Zuordnung zu Antwortkategorien) vergeben.

Eine ausführliche Beschreibung aller Codegruppen mit den jeweils untergeordneten Codeziffern findet sich in Anhang 1.

2.2.2.2. Beispiel eines Dialogs zwischen Interviewer und Befragtem bei einer standardisierten Frage mit entsprechender Codevergabe (Pretestinterview, "live"-Situation)

Fragetext im Fragebogen:

INT.: LISTE P VORLEGEN.
(Liste P zeigt Skala von 0 - 10 mit verbalisierten Endpunkten)
Wie hoch sind Sie zur Zeit durch Berufs- und Hausarbeit alles in allem belastet? Bitte geben Sie Ihre Belastung auf einer Skala zwischen '0' und '10' an. '0' bedeutet, daß Sie 'keine Belastung' fühlen, '10' bedeutet Ihre 'oberste Belastungsgrenze'. Die Werte dazwischen dienen der Abstufung.

Interviewer:

(1. Aktivität)
Wie hoch sind Sie zur Zeit durch Berufs- und Hausarbeit alles in allem belastet? Bitte geben Sie Ihre Belastung auf einer Skala zwischen '0' und '10' an.
Code 22: Interviewer verändert Fragetext erheblich, läßt Schlüsselworte weg, d.h. im vorliegenden Fall Erklärung der Skala.
Code 64: Falsche technische Vorgehensweise (Listenvorgabe nicht beachtet).

Befragter:

.....
(sagt nichts, Interviewer gibt zu wenig Zeit zum Nachdenken und fährt gleich fort)

Interviewer:

(2. Aktivität)
Wie hoch sind Sie zur Zeit durch Berufs- und Hausarbeit alles in allem belastet?
Code 47: Interviewer läßt Befragtem zu wenig Zeit zum Nachdenken.
Code 42: Interviewer wiederholt Frage nicht korrekt, d.h. er liest Frage bei nochmaligem Vorgeben nicht so vor, wie sie im Fragebogen steht.

Befragter:

Ach Gott, das ist aber sehr schwer. Ich meine, es gibt bestimmt viele, die das total ...
(Interviewer fällt ihm ins Wort, beide sprechen; Befragter ist nicht mehr zu verstehen)

- Interviewer: (3. Aktivität)
Liste P, gucken Sie mal die Liste P, 0, P -
wo haben wir denn P?
Code 35: Interviewer klärt technische
Vorgehensweise korrekt.
Code 62: Interviewer hat Befragten unterbrochen.
- Befragter: Ja, aber auf der anderen Seite soll es ja
meinem subjektiven Empfinden entsprechen ...
(Interviewer unterbricht)
- Interviewer: (4. Aktivität)
Ja, richtig.
Code 62: Interviewer unterbricht.
- Befragter: ... und nicht im Vergleich mit anderen Personen.
- Interviewer: (5. Aktivität)
Ja, genau, ganz Ihr subjektives Empfinden.
Code 58: Interviewer macht zulässige Bemerkung
- Befragter: Sieben.
(Befragter antwortet mit Punkt 7 auf der Liste)

Die bei dieser Frage vergebenen Codes für den Interviewer sind demnach: 22, 64, 47, 42, 35, 62, 62, 58.

2.3. Zielsetzung der Untersuchung 1

Mit dem Ziel, ein Verfahren wie das Interaction-Coding-System für die Bewertung von Interviewerverhalten zu erproben, war gleichzeitig die Aufgabe verbunden, die Leistung des ZUMA-eigenen Interviewerstabes zu analysieren.

Wie bereits eingangs erwähnt ermöglicht das Codesystem durch die ihm zugrundeliegenden allgemein gültigen Verhaltensregeln eine Bewertung nach weitgehend objektiven Kriterien. Somit ist die Voraussetzung für die Vergleichbarkeit der Daten, d.h. im konkreten Fall der Interviewerleistung geschaffen. Es kann verglichen werden:

- die Leistung pro Interviewer in einzelnen Verhaltensbereichen,
- die Gesamtleistung pro Interviewer über alle Verhaltensbereiche,
- die Gesamtleistung über alle Interviewer pro Bereich sowie über alle Bereiche hinweg.

Durch diesen Einblick in Einzel- und Gesamtleistung ergeben sich folgende Möglichkeiten:

- Feststellen des Leistungsstandes des Gesamtstabes bei einer Studie,
- eine in systematischer Form gegebene Rückmeldung an den Interviewer über sein Verhalten
 - o pro Interview,
 - o für alle von ihm realisierten Interviews,
 - o über das Gesamtverhalten aller Interviewer in einer Studie, was dem einzelnen die Möglichkeit bietet, den Stellenwert seiner eigenen Leistung im Vergleich zu der anderer zu sehen (Leistungsmotivation),
- gezieltes Einsetzen von Schulungsmaßnahmen in allen im Codesystem charakterisierten Verhaltensbereichen.

In die Verhaltensbewertung bei der Interviewertätigkeit wurden alle zum damaligen Zeitpunkt vorhandenen Mitarbeiter des ZUMA-Interviewerstabes mitbezogen. Es handelte sich dabei um geschulte Interviewer, die alle ein mehrstufiges Grundschulungsprogramm durchlaufen hatten. Für die erforderlichen Qualitätskontrollen und die weiteren Schulungsprozeduren für diese Interviewer schien die Anwendung des Interaction-Coding-Systems aufgrund seiner aufgezeigten Eigenschaften und den sich daraus ergebenden Möglichkeiten eine geeignete Vorgehensweise.

2.4. Beschreibung der durchgeführten Feld- und Schulungsarbeiten

2.4.1. Technische Durchführung der Feldarbeit

Die Technik wurde an zwei Pretestfeldern erprobt, die im Abstand von 6 Monaten durchgeführt wurden. Beteiligt waren pro Studie jeweils 12 Interviewer, von denen jeder 3 Quoteninterviews realisierte und auf Band aufzeichnete. Insgesamt lagen also 72 Interviews vor. Alle Interviewer der ersten Pretest-Studie nahmen auch an der 2. Pretest-Studie teil. In beiden Studien kamen - der Konzeption des Codesystems entsprechend - standardisierte Fragebogen zum Einsatz, die sich in ihren Strukturmerkmalen glichen, d.h. überwiegend aus geschlossenen Fragen mit zum Teil offenen Nachfaßfragen bestanden. Die durchschnittlichen Erhebungszeiten betrugen in beiden Feldern ca. 35-40 Minuten.

2.4.2. Feedback- und Schulungsphase

Nach Abschluß der Feldarbeit zur 1. Studie wurden die Tonbandprotokolle abgehört und für jede Frage die Verhaltensweise jedes Interviewers mit einer Codeziffer bewertet. Für jeden Interviewer wurden pro Interview und pro

Frage die Codeziffern in einem Schema notiert und auf einem Ergebnisblatt zusammengefaßt. Dieses Ergebnisblatt informierte den Interviewer schriftlich auf standardisierte Weise für die von ihm realisierten Fälle über seine angemessenen und unangemessenen Verhaltensweisen und enthielt in Form eines Gesamtscores die Bewertung seiner Leistung pro Verhaltensbereich (vgl. Tabelle 1, Verhaltensbereiche I-IV) und über alle Verhaltensbereiche hinweg. Somit konnte jeder Interviewer seine Leistung pro Interview und für alle Interviews insgesamt ansehen und einen Vergleich der eigenen Leistung mit der anderer ziehen.

Die Reaktion der Interviewer auf die Anwendung einer solchen Technik war durchaus positiv. Sie begrüßten die Möglichkeit, ein Feedback über ihren eigenen Leistungsstand zu erhalten und mit dem des Gesamtstabes vergleichen zu können.

Im Anschluß an diese schriftliche Feedback-Phase wurden gezielte Schulungsmaßnahmen in Einzelsitzungen unternommen und mit jedem Interviewer eines seiner Bänder, in der Regel das fehlerhafteste, abgehört und die aufgetretenen falschen und richtigen Verhaltensweisen im einzelnen besprochen. Die Analyse der selbst erlebten Interviewsituation in bezug auf richtiges und falsches Verhalten erzielte beim einzelnen Interviewer einen stärkeren Lerneffekt und weckte mehr Interesse, als dies bei eher theoretisch gehaltenen Schulungsprozeduren bisher der Fall war. Wie in den Ergebnissen noch zu zeigen sein wird, wirkten sich die gezielten Schulungsmaßnahmen in der nachfolgenden Preteststudie positiv aus.

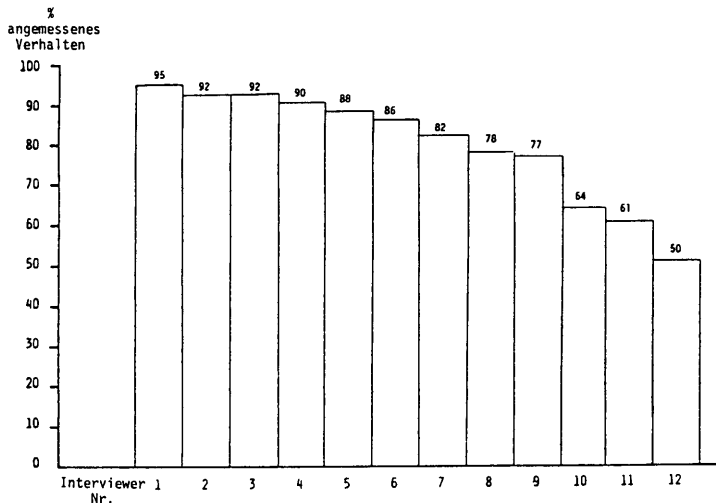
Zur Feedback-Phase der 2. Pretest-Studie sei erwähnt, daß hier die Einzelsitzungen entfielen und der Interviewer lediglich eine Rückmeldung über sein Verhalten in Form eines Ergebnisblattes bekam.

2.5. Ergebnisse zum Interviewerverhalten

Nach der Auswertung der Interviewerverhaltensweisen aus der 1. Erhebung unter Anwendung der beschriebenen Technik konnte für jeden der 12 Interviewer ein Gesamtscore ermittelt werden, der den prozentualen Anteil an angemessenen Verhaltensweisen über alle von ihm realisierten Interviews ausdrückte. Ein Wert von 90 bedeutet also, daß 90% aller für den betreffenden Interviewer registrierten Verhaltensweisen als angemessen beurteilt wurden.

In der 1. Studie streuten die "Interviewerleistungen" beträchtlich, und zwar von 95% für den besten bis zu 50% für den schlechtesten Interviewer (vgl. Abbildung 1).

Abbildung 1: Individuelle Interviewerleistungen (1. Studie)
(prozentualer Anteil angemessener Verhaltensweisen)



Ein Vergleich dieser Werte zwischen den einzelnen Interviewern ist nur insofern sinnvoll, als lediglich die prozentualen Anteile an angemessenen Verhaltensweisen gegenübergestellt werden. Eine darüber hinausgehende Interpretation der Leistungswerte als Maß für die Leistungsqualität eines Interviewers erscheint problematisch, da bei der Codevergabe generell keine Abstufungen bezüglich der "Qualität" einer Verhaltensweise vorgesehen sind. Eine Verhaltensweise kann nur als "richtig" oder "falsch" vercodet werden, eine Differenzierung zwischen diesen beiden Polen ist nicht möglich. Ein folgenschwerer Fehler des Interviewers (z.B. Filterfehler) geht mit gleichem Gewicht - mit der Maßzahl 1 - wie ein Flüchtigkeitsfehler (z.B. "Interviewer macht überflüssige Bemerkung") in die Berechnung des sog. Leistungswertes ein. So ist es durchaus denkbar, daß ein Interviewer, der die Angewohnheit besitzt, häufig überflüssige Bemerkungen zu machen, einen niedrigeren Wert erzielt als ein anderer Interviewer, der nur wenige, dafür aber schwerwiegendere Fehler begeht. Dieser Punkt sollte in einem zur Be-

wertung von Interviewerverhalten konzipierten Codesystem berücksichtigt werden, z.B. durch die Möglichkeit, qualitative Differenzierungen bei der Beurteilung einzelner Intervieweraktivitäten vornehmen zu können.

Für die weitere Analyse des Interviewerverhaltens stand die Klärung zweier Fragen im Vordergrund:

1. Wie verteilen sich angemessene und unangemessene Intervieweraktivitäten auf die einzelnen Bereiche?
2. Gibt es typische Fehlerverteilungen auf einzelne Bereiche für "gute" und "schlechte" Interviewer?

Tabelle 2 zeigt die Verteilung von angemessenen und unangemessenen Intervieweraktivitäten auf die einzelnen Bereiche auf der Grundlage von 35 Interviews der 1. Studie, wobei insgesamt 1415 Aktivitäten vercodet wurden.

Tabelle 2: Verteilung von angemessenen und unangemessenen Intervieweraktivitäten auf die einzelnen Bereiche (1. Studie)

	angemessene Aktivitäten	unangemessene Aktivitäten
Bereich I "Fragestellung"	52.8%	3.6%
Bereich II "Klärung/Nachfrage"	15.6%	9.4%
Bereich III "Anderes Verhalten"	2.0%	11.2%
Bereich IV "Filter"	<u>4.8%</u>	<u>0.6%</u>
Gesamt	75.2%	24.8%
	100.0%	

Die Tabelle zeigt zunächst, daß von sämtlichen Intervieweraktivitäten rund 3/4 (75.2%) als angemessen und 1/4 (24.8%) als unangemessen bewertet wurden. Weiterhin läßt sich der Tabelle entnehmen, daß sich die unangemessenen Verhaltensweisen auf die Bereiche II "Klärung/Nachfrage" und III "Anderes Verhalten" konzentrieren. Besonders auffallend sind die Werte im Bereich III, wo weitaus mehr Fehler als korrektes Verhalten registriert wurden. Dies liegt vor allem daran, daß in diesem Bereich mehr unangemessene als angemessene Verhaltensweisen im Codeschema zu verzeichnen sind (11:2). Eine Verhaltensweise, die als dem Bereich III zugehörig klassifiziert wird, be-

sitzt damit von vornherein eine hohe Wahrscheinlichkeit, auch als unangemessen eingestuft zu werden.

Tabelle 3: Vergleich der Fehlerstruktur zwischen der Gruppe der drei besten und der drei schlechtesten Interviewer anhand des durchschnittlichen Fehleranteils pro Bereich (1. Pretest-Studie)

	<u>BESTE GRUPPE</u>		<u>SCHLECHTESTE GRUPPE</u>	
	angemessenes Verhalten	unangemessenes Verhalten	angemessenes Verhalten	unangemessenes Verhalten
Bereich I "Fragestellung"	100.0%	0.0%	81.7%	18.3%
Bereich II "Klärung/Nachfrage"	90.7%	9.3%	43.0%	57.0%
Bereich III "Anderes Verhalten"	13.0%	87.0%	2.3%	97.7%
Bereich IV "Filter"	100.0%	0.0%	90.7%	9.3%

Bei einem Vergleich der Verteilung unangemessener Verhaltensweisen zwischen der Gruppe der drei besten Interviewer und der Gruppe der drei schlechtesten Interviewer (vgl. Tabelle 3) sind durchaus gruppenspezifische Fehlerverteilungen zu erkennen. Die "beste" Interviewergruppe ist gekennzeichnet durch absolute Fehlerfreiheit in den Bereichen I und IV, eine relativ geringe Fehlerquote im Bereich II sowie eine hohe Fehlerquote im Bereich III, die jedoch, wie bereits erwähnt, weitgehend systembedingt ist. Arbeitet der "gute" Interviewer also in den Bereichen, in denen sog. "scripted rules" (wie z.B. Vorlesen des Fragetextes) zu befolgen sind (Bereiche I und IV), fehlerfrei, so zeigen die Interviewer der "schlechtesten" Gruppe hier bereits unangemessene Verhaltensweisen. Am deutlichsten ist jedoch der Unterschied zwischen beiden Gruppen im Bereich II "Klärung/Nachfrage", wo in der "besten" Gruppe nur 9.3% der Verhaltensweisen unangemessen waren, in der "schlechtesten" Gruppe dagegen mehr als die Hälfte, nämlich 57.0%. Diese Werte machen deutlich, daß anschließende korrigierende Schulungsmaßnahmen besonders auf die Verbesserung von Verhaltensweisen aus dem Bereich "Klärung/Nachfrage" gerichtet sein mußten.

ZUMA

Die Ergebnisse der zweiten Studie sollten zeigen, ob und in welchem Umfang die gezielten Schulungsmaßnahmen zu Leistungsverbesserungen bei den beteiligten Interviewern geführt hatten.

Abbildung 2: Individuelle Interviewerleistungen (2. Studie)
(prozentualer Anteil angemessener Verhaltensweisen)

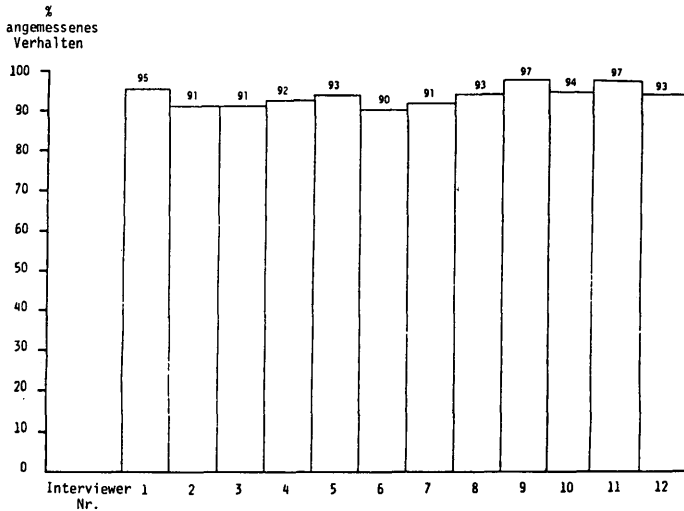


Abbildung 2 zeigt, daß die Variationsbreite der individuellen Interviewerleistungen sich stark reduzierte, was auf eine deutliche Leistungsverbesserung insbesondere bei schlechten Interviewern schließen läßt. Dieses darf jedoch nur bedingt als echte Qualitätsverbesserung interpretiert werden, da sich beim Abhören der Bänder zur zweiten Studie zeigte, daß die Interviewer in Kenntnis der Bewertungssituation eine Fehlervermeidungstendenz entwickelten, d.h. sich bedeutend "vorsichtiger" in ihrer Vorgehensweise verhielten. Fast alle Interviewer waren offensichtlich bemüht, von vornherein Situationen zu vermeiden, in denen Verhaltensweisen abverlangt wurden, die potentiell die Gefahr in sich bergen, als unangemessen klassifiziert zu werden. Das Bewußtsein um die Bewertung ihres Verhaltens mit Hilfe des Codesystems ließ sie "formaler" vorgehen, als dies bei der ersten Studie der Fall war.

ZUMA

Ebenso ist ein direkter Vergleich dieser Werte zwischen erster und zweiter Studie durch die Tatsache eingeschränkt, daß in der zweiten Studie ein strukturell ähnlicher, jedoch inhaltlich anderer Fragebogen zur Anwendung kam und daß die in der zweiten Studie rekrutierten Zielpersonen mit denen der ersten nicht identisch waren, also potentiell andere Reaktionsweisen zeigen konnten, was sich insbesondere auf die Verhaltensweisen des Interviewers bei der Klärung (im Bereich II) auswirken konnte. Ähnliches gilt für die Verteilung von angemessenen und unangemessenen Intervieweraktivitäten in bestimmten Bereichen (vgl. Tabelle 4).

Tabelle 4: Verteilung von angemessenen und unangemessenen Intervieweraktivitäten auf die einzelnen Bereiche (2. Studie)

	<u>angemessene Aktivitäten</u>	<u>unangemessene Aktivitäten</u>
Bereich I "Fragestellung"	71.4%	2.8%
Bereich II "Klärung/Nachfrage"	12.2%	1.1%
Bereich III "Anderes Verhalten"	2.3%	2.8%
Bereich IV "Filter"	<u>6.8%</u>	<u>0.6%</u>
Gesamt	92.7%	7.3%

Ein Vergleich mit den Werten der ersten Studie (vgl. Tabelle 2) zeigt, daß die Fehlerquote besonders in den Bereichen II und III merklich geringer geworden ist. Die Interpretation der Werte in Richtung Qualitätsverbesserung ist - wie bereits erwähnt - eingeschränkt zu sehen, als gerade in diesen Bereichen die Interviewer die Möglichkeit haben, gezielt Aktivitäten einzuschränken, um potentielle Fehler zu vermeiden (z.B. Minimum an Klärungsaktivitäten, Unterlassen von zusätzlichen Bemerkungen usw.), was in der zweiten Studie der Fall war.

Eine tatsächliche Qualitätsverbesserung liegt in den Bereichen I und IV (Fragestellung und Filter), beides Bereiche, die sowohl fragebogen- als auch befragtenunabhängig sind, und in denen der Interviewer durch die dort vorgeschriebenen festen Regeln wenig Beeinflussungsmöglichkeiten hat.

Ganz besonders deutlich drücken sich die Verhaltensänderungen gegenüber der ersten Studie in den Werten der drei "schlechtesten" Interviewer (Nr. 10,

11 und 12 in den Abbildungen 1 und 2) aus, wenn man die einzelnen Bereiche betrachtet (vgl. Tabelle 5).

Tabelle 5: Vergleich der Fehlerstruktur zwischen der Gruppe der drei besten und der drei schlechtesten Interviewer anhand des durchschnittlichen Fehleranteils pro Bereich (2. Studie)

	<u>BESTE GRUPPE</u>		<u>SCHLECHTESTE GRUPPE</u>	
	<u>angemessenes Verhalten</u>	<u>unangemessenes Verhalten</u>	<u>angemessenes Verhalten</u>	<u>unangemessenes Verhalten</u>
Bereich I "Fragestellung"	95.7%	4.3%	94.7%	5.3%
Bereich II "Klärung/Nachfrage"	90.3%	9.7%	90.0%	10.0%
Bereich III "Anderes Verhalten"	18.0%	82.0%	75.3%	24.7%
Bereich IV "Filter"	90.7%	9.3%	100.0%	0.0%

Ein Vergleich mit den entsprechenden Werten der ersten Studie zeigt bei der "schlechtesten" Gruppe in allen Bereichen drastische Verringerungen der Fehleranteile. Auch hier gilt, daß die Abnahme der Fehlerquote in den Bereichen II und III aus den erwähnten Gründen nicht nur als Qualitätsverbesserung der Interviewerleistungen interpretiert werden darf, während man bei den verringerten Fehlerhäufigkeiten in den Bereichen I und IV durchaus von positiven Schulungseffekten ausgehen kann.

Ein Vergleich beider Gruppen verdeutlicht die Folgen des überängstlichen Fehlervermeidungsverhaltens der "schlechten" Interviewer. Im Bereich III "Anderes Verhalten" erzielte die Gruppe sogar Fehlerquoten, die weit unter denen der "besten" Gruppe liegen. Vergleicht man die vorliegenden Werte der "besten" Gruppe mit denen der ersten Studie, dann fällt auf, daß die Fehleranteile in den Bereichen II und III relativ konstant geblieben sind und in den Bereichen I und IV sogar angestiegen sind. Dies könnte als sog. "Lorbeerereffekt" interpretiert werden: Die Kenntnis der eigenen guten Leistungen in der ersten Studie führte nun zu Nachlässigkeit und Unkonzentriertheit.

Bei einem Vergleich der ersten mit der zweiten Studie bleiben neben der eingeschränkten Interpretationsmöglichkeit der rein quantitativen Interviewerleistungswerte und den Fehlerquellen noch einige Aspekte, die als Verhaltensänderungen als "reine" Folge der Schulungsmaßnahmen gesehen werden können. Wie bereits erwähnt, finden sich diese Aspekte bei denjenigen Interviewertätigkeiten, die sowohl befragten- als auch fragebogenunabhängig sind. Dies gilt weitgehend für den Bereich "Fragestellung", bei dem es um das Vorlesen des Fragetextes geht. Ein Vergleich der quantitativen Ergebnisse (vgl. Tabellen 2 und 4) zeigt in diesem Bereich zwar keine gravierenden Veränderungen des Verhältnisses von angemessenen und unangemessenen Verhaltensweisen; dennoch konnte durch Betrachtung einzelner Codeziffern in diesem Bereich eine qualitative Verbesserung gegenüber der ersten Studie insofern festgestellt werden, als nämlich die Tendenz der Interviewer, den Fragetext geringfügig verändert vorzulesen (Codeziffer 12, zählt als angemessenes Verhalten), merklich nachließ. Stattdessen verstärkte sich die Tendenz, den Fragetext völlig korrekt vorzulesen (Codeziffer 11).

Auch innerhalb des Bereiches "Anderes Verhalten", und zwar speziell bei der Codegruppe 60, die im wesentlichen befragten- und fragebogenunabhängige Aktivitäten beschreibt, konnten in der zweiten Studie positive Verhaltensänderungen beobachtet werden, die darin bestanden, die in der ersten Studie relativ häufig registrierten unangemessenen Aktivitäten (wie z.B. Code 63 "Interviewer gibt persönliche Meinung oder Bewertung ab" oder Code 66 "Interviewer formuliert Antwort des Befragten noch einmal grundlos mit anderen Worten") nun zu vermeiden. Dieser Lerneffekt trat besonders stark bei den leistungsschwächeren Interviewern auf.

Generell gibt es Hinweise dafür, daß das Lernziel, gewohnheitsmäßig angelegene, überflüssige Aktivitäten abzubauen, vom Interviewer relativ schnell und problemlos erreicht werden kann, schneller als z.B. das Erlernen adäquater, flexibler Aktivitäten als Reaktion auf bestimmte Verhaltensweisen des Befragten.

Weiterhin gibt es Hinweise dafür, daß Regeln bzw. Aufgaben, die der Interviewer dem Fragebogen entnehmen kann (sog. "scripted rules"), wie z.B. Fra-

getext, Antwortalternativen, Intervieweranweisungen und Filterhinweise, einfacher erlern- und anwendbar sind als diejenigen Aufgaben, die der Interviewer durch Schulungs- und Trainingsmaßnahmen gelernt haben sollte und je nach Interviewsituation flexibel anwenden muß. In diesem Bereich sind verstärkte Schulungsmaßnahmen erforderlich.

2.6. Beurteilung der Interaction-Coding-Technik zur Bewertung von Interviewerverhalten

Für den hier dargestellten Anwendungsbereich, die Analyse von Interviewerverhalten, ist die Technik des Interaction-Coding anwendbar. Wie sich gezeigt hat, kann die Gesamtleistung jedes Interviewers in einer Studie weitgehend objektiv beurteilt werden, wobei auch eine Diagnose der Interviewerleistung in einzelnen Verhaltensbereichen möglich ist. Damit bietet das Verfahren den Vorteil, Schulungsmaßnahmen gezielt bereichs- und personenspezifisch aufgrund von weitgehend objektiv durchgeführten Qualitätskontrollen einsetzen zu können. Gleichzeitig führen die Einzeldiagnosen zur Feststellung der Gesamtleistung eines Interviewerstabes.

Auch wenn man die erwähnten besonderen Umstände berücksichtigt, unter denen die Interviews der vorne beschriebenen zweiten Studie durchgeführt wurden, läßt sich sagen, daß sich die mit Hilfe der Interaction-Coding-Technik aufgedeckten Schwachstellen in der Interviewerleistung besonders bei den leistungsschwächeren Interviewern beheben lassen, d.h. daß sich Leistungssteigerungen erzielen lassen. Dies führt zur Verringerung von Interviewereffekten in der Erhebung, ein für die Optimierung des Datenerhebungsprozesses wichtiger Aspekt.

3. Untersuchung 2 - Bewertung der Qualität von Fragen

3.1. Allgemeine Bemerkungen zum Bewertungssystem

Der Versuch, mit Hilfe der Interaction-Coding-Technik etwas über die Güte von Fragen auszusagen, liegt in der Verfahrensweise mit dem System selbst begründet. Wie bereits erwähnt, wird jede verbale Aktivität, d.h. eine im Codesystem beschriebene Verhaltensweise, durch eine Codeziffer erfaßt. Je mehr Aktivitäten also bei einer Frage stattfinden, desto mehr Codes werden vergeben. Man kann davon ausgehen, daß dies eher bei unklar operationalisierten Fragetexten der Fall ist, die Rückfragen durch den Befragten und somit Zusatzklärungen durch den Interviewer provozieren, als bei Fragen mit eindeutig gesetztem Stimulus, die ohne zusätzliche Aktivitäten, d.h.

Rückfragen bzw. deren Klärung, beantwortet werden können. So kann die Häufigkeit der Vergabe von Codes, bedingt durch die Länge der Interaktion zwischen Befragtem und Interviewer, ein Indiz für die Funktionsfähigkeit einer Frage sein (vgl. hierzu auch die Arbeiten von MORTON-WILLIAMS, 1983, und CANNELL, 1971).

Der übliche Weg, Fragen auf ihre Funktionsfähigkeit hin zu testen, ist die Durchführung von Pretests. Dabei nimmt der Interviewer eine wichtige Funktion insofern ein, als er in sogenannten Erfahrungsberichten seine Beobachtungen zum Funktionieren einzelner Fragen niederschreibt. Diese Interviewerberichte sind von der subjektiven Wahrnehmung des einzelnen Interviewers geprägt und beschränken sich meist nur auf die für den Interviewer schwerwiegenden Probleme, die sich in seinen Interviews ergaben. Ein objektiver und realitätscher Eindruck von der Interviewsituation ist auf diesem Wege nicht möglich. Da mit der Interaction-Coding-Technik der gesamte Interviewverlauf systematisch registriert wird, lag es nahe, dieses Verfahren auf seine Brauchbarkeit hinsichtlich der Bestimmung der Qualität von Fragen zu testen.

3.2. Kurzbeschreibung der Technik mit Beispielen

Für den Einsatz der Interaction-Coding-Technik zur Qualitätsbewertung von Fragen wurde von den Autoren ein neues Codesystem entwickelt, das für das standardisierte mündliche Interview konzipiert ist. Das System basiert in seiner Anwendung auf denselben Grundprinzipien wie das in Untersuchung 1 bereits ausführlich beschriebene, unterscheidet sich jedoch von diesem in zwei wesentlichen Punkten:

1. Es enthält zusätzlich neben dem Katalog aller Interviewerverhaltensweisen eine Aufzählung aller im Interview möglichen Verhaltensweisen des Befragten.
2. Sein formaler Aufbau unterscheidet nicht zwischen angemessenen und unangemessenen Verhaltensweisen.

ad 1. Bereits in dem Codesystem zur Interviewerverhaltensbewertung in Untersuchung 1 erwies sich das Fehlen der Registrierung der Befragtenreaktionen als Mangel, da Interviewerverhalten zum großen Teil vom Befragtenverhalten abhängig ist und in manchen Fällen nur durch dieses erklärbar wird. Für die Bewertung der Qualität einer Frage, gemessen an dem Ausmaß von In-

teraktion zwischen Interviewer und Befragtem, ist die Registrierung der Befragtenaktivitäten unerlässlich.

ad 2. Setzt man die Interaction-Coding-Technik zur Bewertung der Fragenqualität ein, so geht es darum, die Qualität einer Frage über die formalen Eigenschaften der Verhaltensweisen bzw. Interaktionen zu erschließen und nicht über deren Bewertung. Aus diesem Grund wird in dem für die vorliegende Untersuchung konstruierten Codeschema nicht nach angemessenen und unangemessenen Verhaltensweisen differenziert. Würde ein solches Codeschema auf der Bewertung der Verhaltensweisen aufbauen, so würde zusätzlich zur Bewertung der Qualität von Fragen noch eine subjektiv geprägte Bewertungsdimension Einfluß gewinnen.

Die folgenden Erläuterungen zum Codeschema konzentrieren sich nicht mehr auf dessen prinzipielle Struktur und Anwendungsweise. Diese wurden bei der Abhandlung zur Untersuchung 1 bereits ausführlich beschrieben. Es soll hier lediglich der formale Aufbau vorgestellt werden. Das Codeschema besteht aus einer Aufzählung möglichst aller Verhaltensweisen, die der Interviewer und der Befragte während des Interviews zeigen können, wie z.B.: Vorlesen des Fragetextes, Verständnisfragen des Befragten, Klärungen der Rückfragen durch den Interviewer, Antwort des Befragten.

- Komplex 1 umfaßt alle Verhaltensbeschreibungen, die die Stimuluspräsentation durch den Interviewer, d.h. das Vorlesen der Frage bzw. Antwortkategorien, Skalen oder Items repräsentieren und die vor der ersten Reaktion des Befragten stattfinden. Die Codeziffern für diese Verhaltensweisen liegen in der Codegruppe 100.
- Komplex 2 beschreibt alle Verhaltensweisen des Befragten. Die Codeziffern hierfür liegen in der Codegruppe 200.
- Komplex 3 beschreibt alle über das Vorlesen der Frage hinausgehenden Interviewerverhaltensweisen, wie z.B. Klärungen oder Nachfragen. Es handelt sich dabei um Verhaltensweisen nach der ersten Reaktion des Befragten. Sie werden durch die Codegruppe 300 repräsentiert.
- Komplex 4 befaßt sich mit der Registrierung der Antworten durch den Interviewer im Fragebogen und wird in der Codegruppe 400 erfaßt.

Eine detaillierte Beschreibung des Codeschemas findet sich in Anhang 2.

3.2.1. Vergabe von Codes

Jeder der auf Band registrierten Verhaltensweisen – sowohl des Interviewers als auch des Befragten – wird eine entsprechende Codeziffer zugeordnet. Dabei werden die Codeziffern Frage für Frage notiert, wobei die Anzahl der pro Frage vergebenen Codes von der Länge des Interaktionsprozesses zwischen Interviewer und Befragtem abhängt. Zwei Beispiele für mögliche Interaktionsverläufe zwischen Interviewer und Befragtem sollen die Anwendung des Codeschemas verdeutlichen.

Beispiel 1 zeigt einen idealen Verlauf bei der Beantwortung einer Frage, d.h. der Interviewer liest den Fragetext und die Antwortalternativen wörtlich vor, der Befragte gibt eine den Antwortkategorien entsprechende Antwort ohne Rückfragen oder Zwischenäußerungen:

Fragetext im Fragebogen Beginnen wir mit einigen Fragen zur wirtschaftlichen Lage: Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in der Bundesrepublik:

INT.: ANWORTVORGABEN VORLESEN.

- sehr gut
- gut
- teils gut/teils schlecht
- schlecht oder
- sehr schlecht?

Interviewer Code 111: liest Fragetext wörtlich vor (wie oben dargestellt).

Code 121: liest laut Anweisung die Antwortvorgaben wörtlich vor (wie oben dargestellt).

Befragter Teils gut, teils schlecht.

Code 201: Befragter gibt adäquate Antwort bei einer geschlossenen Frage.

Die Codes werden in der Reihenfolge ihres Auftretens notiert: 111,121,201.

Beispiel 2 zeigt alternativ dazu einen anderen möglichen Verlauf bei der Beantwortung derselben Frage:

Fragetext im Fragebogen Beginnen wir mit einigen Fragen zur wirtschaftlichen Lage: Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in der Bundesrepublik:

INT.: ANWORTVORGABEN VORLESEN.

- sehr gut
- gut
- teils gut/teils schlecht

ZUMA

- schlecht oder
- sehr schlecht?

Interviewer Beginnen wir mit einigen Fragen zur wirtschaftlichen Lage: Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in der BRD:

- sehr gut
- gut
- teils gut/teils schlecht
- schlecht oder
- sehr schlecht?

Code 112: Interviewer liest Fragetext geringfügig verändert vor.

Code 121: Interviewer liest Antwortvorgaben wörtlich vor.

Befragter Ziemlich mies, würde ich sagen.

Code 222: Befragter gibt inadäquate Antwort (entspricht nicht den Antwortvorgaben der geschlossenen Frage).

Interviewer Bitte entscheiden Sie sich für eine der vorgegebenen Möglichkeiten. Ich lese Sie Ihnen noch mal vor: sehr gut, gut, teils gut/teils schlecht, schlecht oder sehr schlecht?

Code 303: Interviewer wiederholt Frage oder Teile davon korrekt.

Befragter Schlecht.

Code 201: Befragter gibt adäquate Antwort bei geschlossener Frage.

Es werden folgende Codes vergeben: 112,121,222,303,201.

3.3. Feldarbeit und Fragebogen

3.3.1 Feldarbeit

Vor der Ergebnisdarstellung der Untersuchung sollen kurz die Durchführung der Feldarbeit und das der Untersuchung zugrundeliegende Instrument angesprochen werden.

Von 12 ZUMA-Interviewern wurden 60 Adresseninterviews realisiert. Die Adressen wurden von den Interviewern auf der Grundlage einer Random-route ermittelt, innerhalb des Haushalts wurde die Zielperson nach einem Zusatzverfahren bestimmt. Die Verteilung der demographischen Merkmale weicht von derjenigen herkömmlicher Bevölkerungsumfragen ab. Da die vorliegende Unter-

suchung eher den Charakter einer Pilotstudie hatte, erhebt sie nicht den Anspruch, in der Zusammensetzung der Stichprobe Repräsentativität widerzuspiegeln. Es wurde lediglich darauf geachtet, daß Alter, Geschlecht und Bildung ausreichend differenziert vertreten waren.

3.3.2. Der Fragebogen

Der Fragebogen bestand aus einer Auswahl von Fragen aus der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1984. Da vorangegangene Erfahrungen gezeigt haben, daß die Durchführung der Vercodungsarbeiten sehr zeitaufwendig ist, mußte die Anzahl der mittels Codesystem zu analysierenden Fragen begrenzt werden. Dabei wurde darauf geachtet, daß in ihrer Komplexität unterschiedliche Fragetypen (Alternativfragen, Fragen mit Verbal-skalen, mit numerischer Skala, Rangreihenbildung, Kärtchen- und Listenvorlagen usw.) miteinbezogen wurden. Anzumerken ist noch, daß bei Fragen mit Itembatterien die einzelnen Items als eigenständige Fragen bei der Ver-codung behandelt wurden.

3.4. Ergebnisdarstellung

3.4.1. Definition des Qualitätsbegriffs

Vor der Betrachtung, inwieweit die angewandte Technik Aussagen zur Qualität von Fragen zuläßt, soll versucht werden, den Begriff der "Qualität" zu definieren. Eine wichtige Voraussetzung für die gute Qualität einer Frage ist dann gegeben, wenn die vom Forscher intendierte Bedeutung mit der vom Befragten wahrgenommenen Bedeutung übereinstimmt. Unter diesem Qualitätsaspekt sind die Bemühungen vieler Forscher zu verstehen, Fragetexte semantisch eindeutig zu formulieren. Auch die Bemühungen, die Bedeutungswahrnehmungen des Befragten nachträglich zu erfragen und damit zu kontrollieren, wie es z.B. BELSON (1982) versucht hat, sind letztlich aus diesem Qualitätsbegriff heraus zu verstehen.

Das Problem besteht jedoch darin, daß eine Überprüfung des Fragenverständnisses in Umfragen mit größerer Stichprobe aus Zeitbeschränkungen nicht möglich ist. Ebenso verbieten in den meisten Fällen finanzielle Gründe eine Nacherhebung der Stimulusbedeutung. Da man einerseits davon ausgehen kann, daß das Verständnis von Fragenstimuli interindividuell stark variiert, man andererseits aber nur sehr unzureichende Möglichkeiten einer Kontrolle hat, sollte man versuchen, alle Faktoren des Interviewprozesses, die die Varia-

bilität des Fragenverständnisses zusätzlich erhöhen könnten, zu vermeiden bzw. unter Kontrolle zu bringen.

Wie eingangs erwähnt kann in erster Linie der Interviewer als Störfaktor in den Beantwortungsprozeß eingreifen, was zu möglichen Verzerrungen in dem vom Forscher intendierten Frageninhalt führen kann. Da die Eingriffe des Interviewers meistens dann erfolgen, wenn Nachfragen durch den Befragten bzw. unzulängliche Antworten vorliegen, gibt die Anzahl dieser Eingriffe selbst schon einen möglichen Indikator für die Formulierungsgüte einer Frage und damit für einen wichtigen Qualitätsaspekt ab.

Natürlich gibt es weitere Möglichkeiten, den Qualitätsbegriff von Fragen durch Operationalisierung näher zu präzisieren. Eine naheliegende Möglichkeit wäre es zu beachten, in welchem Ausmaß eine Frage zu adäquaten Antworten führt, d.h. zu Antworten, die der Fragenintention bzw. den Vorgaben entsprechen, ohne daß irgendwelche Hinweise auf Invalidität vorliegen. So könnte man eine Frage A dann als qualitativ besser als eine Frage B ansehen, wenn A zu einer höheren Anzahl "adäquater" Antworten führt als B. Leider kann man dieser Operationalisierung nicht ohne weiteres vertrauen, da man bei einem Eingriff des Interviewers nicht weiß, ob die von einem Befragten gegebene adäquate Antwort auch ohne diesen Eingriff zustande gekommen wäre. Dieser Zweifel führt wieder zurück zu der Notwendigkeit, auch die Interaktion von Interviewer und Befragtem zu erfassen, die das Codesystem aufgrund seiner strukturellen Beschaffenheit leisten kann.

3.4.2. Ergebnisse zur Qualitätsbestimmung

In die Bewertung mit dem Codesystem wurden 17 Fragen einbezogen, darunter Fragen mit langen Itembatterien. Die einzelnen Items wurden in der Vercodung jeweils als eigenständige Frage behandelt, so daß man der Berechnung nicht ein N=17 zugrundelegen muß, sondern N=57 (Fragen). Da die hier vorgestellte Untersuchung eher den Charakter einer Pilotstudie hat, d.h. der Umgang mit dem Codesystem für diesen Anwendungsbereich erst erprobt werden sollte, basieren die dargestellten Ergebnisse auf einfachen Auswertungsschemata. Ein erster Auswertungsschritt sollte zeigen, ob durch Betrachtung der Häufigkeiten der Codevergabe pro Frage Aussagen über deren Funktionieren bzw. Nichtfunktionieren angestellt werden können. Damit beziehen sich die zugrundegelegten Analysen auf elementare Formen der Interaktion. Diffe-

renzierte Formen der Interaktion, die durch die Betrachtung bestimmter Muster von Codeziffern angegangen werden könnten, müssen späteren Analysen vorbehalten bleiben.

Tabelle 6 zeigt die Verteilung von möglichen Interaktionsverläufen zwischen Interviewer und Befragtem für alle Fragen und alle realisierten Interviews in Prozent.

Tabelle 6: Häufigkeit möglicher Interaktionsverläufe über alle Fragen und alle Interviews in Prozent (N=3346)

I	II	III
Fälle mit Idealverlauf	Fälle mit adäquater Antwort <u>ohne</u> idealen Verlauf	Fälle ohne adäquate Antwort
29.9%	60.5%	9.6%

Die Tabelle unterscheidet zwischen Fällen

- mit idealem Verlauf, d.h.: Der Interviewer liest den Fragetext wörtlich vor, der Befragte antwortet adäquat, ohne daß zusätzliche Aktivitäten zwischen Stimulusdarbietung und Antwort liegen.
- mit adäquater Antwort ohne idealen Verlauf, d.h. zum Beispiel: Der Interviewer liest den Fragetext wörtlich vor, der Befragte stellt eine Verständnisrückfrage, der Interviewer klärt diese korrekt, der Befragte antwortet adäquat, d.h. den Vorgaben bzw. der Fragenintention entsprechend.
- ohne adäquate Antwort, d.h. zum Beispiel: Der Interviewer liest den Fragetext wörtlich vor, der Befragte kann die Frage den Vorgaben entsprechend nicht beantworten (Verständnisschwierigkeiten, die vom Interviewer nicht gelöst werden können, "weiß nicht", "keine Angabe").

Die Verteilung der in Tabelle 6 dargestellten Interaktionsverläufe beruht auf der Definition des Idealverlaufs im "engsten" Sinne. Idealverläufe im engsten Sinne sind solche, bei denen der Interviewer den Fragestimulus Wort für Wort korrekt vorgibt (vgl. Codeschema in Anhang 2, Codeziffer 111). Daneben stehen Idealverläufe im weiteren Sinne, bei denen dem Interviewer geringfügige Abweichungen beim Vorlesen des Fragetextes, der Antwortkategorien oder der Items zugestanden werden. Diese sind Worteinfügungen wie "und", "oder", bzw. kleine Wortveränderungen, die den Sinn der Frage bzw.

deren Bezugsrahmen nicht verändern (vgl. Codeschema in Anhang 2, Codeziffer 112).

Läßt man zu, daß diese Verläufe noch als ideal definiert werden, so verschiebt sich die in Tabelle 6 dargestellte Verteilung folgendermaßen (vgl. Tabelle 7):

Tabelle 7: Häufigkeit möglicher Interaktionsverläufe (im weiteren Sinne) über alle Fragen und alle Interviews in Prozent (N=3346)

I	II	III
Fälle mit Idealverlauf	Fälle mit adäquater Antwort <u>ohne</u> idealen Verlauf	Fälle ohne adäquate Antwort
49.9%	40.5%	9.6%

Daß diese Fälle noch als ideal definiert werden, schien deshalb sinnvoll, da sie sich nur unwesentlich von den streng definierten Idealverläufen unterscheiden und - wie man in Tabelle 7, Spalte II sieht - diejenigen ohne idealen Verlauf um 20% reduzieren. Somit kann man davon ausgehen, daß sich die in Spalte II verbleibenden Fälle eindeutiger von den Idealverläufen abheben. Auf die Fälle in Gruppe III wird später noch ausführlich eingegangen.

Anhand dieser beiden Tabellen wurden die möglichen Interaktionsverläufe über alle Fragen demonstriert. Natürlich ist dies auch für jede einzelne Frage möglich, d.h. für jede Frage läßt sich der Anteil der Fälle mit Idealverlauf, der Fälle mit adäquater Antwort ohne idealen Verlauf sowie der Fälle ohne adäquate Antwort ermitteln. Diese Werte liefern zunächst einmal einen relativ groben Hinweis auf das "Funktionieren" bzw. die Qualität einer Frage. Die dargestellten Ergebnisse zeigen neben der grundsätzlichen Fähigkeit des Codesystems zur Differenzierung zwischen Fallgruppen außerdem, daß zusätzliche Aktivitäten jenseits des Idealverhaltens auch tatsächlich auftreten, eine Erkenntnis, die einerseits nicht besonders überrascht, in ihrem Ausmaß aber erst durch ein geeignetes Codesystem bestimmbar wird.

Ziel der Tabelle 8 ist es darzustellen, wie sich die Zusatzaktivitäten, d.h. die über den Idealverlauf hinausgehenden Aktivitäten, auf die einzel-

nen Fragen verteilen. Zu diesem Zweck wurde für jede Frage die durchschnittliche Anzahl von Zusatzaktivitäten berechnet und in 4 Intervalle aufgeteilt. Verließ die Interaktion bei einer Frage ideal, so ging dieser Idealfall mit 0 Zusatzaktivitäten in die Berechnung ein.

Tabelle 8: Intervalle durchschnittlicher Häufigkeiten von Zusatzaktivitäten pro Interview für alle Fragen

Häufigkeitsintervalle	0.1-0.9	1.0-1.9	2.0-2.9	3.0 und mehr
Anzahl der Fragen (N=57)	35	15	4	3
in Prozent	61.4%	26.3%	7.0%	5.3%

Die Tabelle zeigt, daß bei 61.4% der Fragen (N=57) die durchschnittliche Anzahl der über den Idealverlauf hinausgehenden Zusatzaktivitäten in das Intervall 0.1-0.9 fällt, d.h. daß hier im Durchschnitt der Interviewverlauf nur geringfügig vom Idealverlauf abweicht. Es fällt auf, daß immerhin 38.6% aller Fragen durchschnittliche Zusatzaktivitäten von über 1.0 aufweisen und 12.3% durchschnittliche Zusatzaktivitäten von über 2.0. Die Durchschnittswerte scheinen relativ klein. Es gibt jedoch in jeder Gruppe sog. "Ausreißer"; selbst in der Gruppe mit der geringsten Zusatzaktivität, im Intervall von 0.1-0.9 liegt immerhin ein "Ausreißer" mit 16 Aktivitäten, und im letzten Intervall gibt es Spitzenwerte bis zu 30 Aktivitäten.

In den bisherigen Ausführungen dieses Abschnitts standen vor allem jene Aktivitäten im Vordergrund, die zwischen Stimulusdarbietung und adäquater Befragtenreaktion lagen und die über alle Fragen dargestellt waren. Im folgenden sollen Interaktionsverläufe an einigen Einzelfragen vorgestellt werden. Dabei sind besonders die in Tabelle 6 und 7 definierten Fälle interessant, die inadäquates Antwortverhalten zeigten. Der Prozentsatz dieser Gruppe (9.6%, vgl. Tabelle 7) erweist sich in der vorliegenden Untersuchung erheblich höher als der bei Survey-Umfragen normalerweise festgestellte Prozentsatz der missing values. Dies liegt vor allem daran, daß sich diese Gruppe bei der Bewertung durch das Codesystem durchaus nicht als homogen erwies und interessante Details verbarg. Diese sollen im folgenden bei Interaktionsverläufen zu bestimmten Fragen aufgezeigt werden (vgl. Abbildung 3).

ZUMA

Abbildung 3: Auszug aus Frage 9, Items A, E und G

INT.: graues Kärtchenspiel mischen und übergeben

Nun zu einem anderen Bereich: Ich habe hier einige Meinungen über Staat und Wirtschaft der Bundesrepublik. Sagen Sie mir bitte zu jeder Meinung, ob Sie ihr voll zustimmen, eher zustimmen, eher nicht zustimmen oder überhaupt nicht zustimmen.

INT.: zu jeder Aussage eine Antwortziffer einkreisen

stimme voll zu	stimme eher zu	stimme eher nicht zu	stimme überhaupt nicht zu	weiß nicht
1	2	3	4	8

- A In unserer Gesellschaft muß jeder für sich schauen, daß er auf einen grünen Zweig kommt. Es hilft nicht viel, sich mit anderen zusammenzuschließen, um politisch oder gewerkschaftlich für seine Sache zu kämpfen.
- E Wenn die Leistungen der sozialen Sicherung, wie Lohnfortzahlungen im Krankheitsfall, Arbeitslosenunterstützung und Frührenten, so hoch sind wie jetzt, führt dies nur dazu, daß die Leute nicht mehr arbeiten wollen.
- G Die wirtschaftlichen Gewinne werden heute in der Bundesrepublik im großen und ganzen gerecht verteilt.

Erläuterung zur Technik dieser Frage:

Die einzelnen Karten enthalten jeweils den Formulierungstext des Items (optische Vorlage); die Skalenpunkte 1 bis 4 werden dem Befragten mittels Fragetext verbal vermittelt (akustische Vorgabe).

Tabelle 9: Prozentsätze der Interaktionsverläufe für 3 Items zu Frage 9

Frage	I Ideal- verläufe	II Fälle mit adä- quater Antwort <u>ohne</u> idealen Verlauf	III Fälle mit inadäquaten Antworten		Gesamt
			missings (KA,wn,vw)*	inadäquate Antwort, im Fragebogen adäquat bezeichnet	
9A	15.0%	46.7%	3.4%	34.9%	38.3%
9E	20.0%	46.7%	3.3%	30.0%	33.3%
9G	16.7%	56.6%	-	26.7%	26.7%

*KA = Keine Angabe im Fragebogen

wn = weiß nicht

vw = verweigert

Betrachtet man in Tabelle 9 die Gruppe III mit inadäquatem Antwortverhalten, so sieht man, daß sich hier 2 Untergruppen bilden:

1. die Gruppe der sogenannten "missing values" mit den Varianten "keine Angabe" (d.h. im Fragebogen fehlt jegliche Eintragung), "weiß nicht" und "verweigert";
2. Fälle, die im Fragebogen als adäquate Antwort verzeichnet sind, sich jedoch beim Vercodungsprozeß als inadäquat erwiesen haben.

ad 1. Diese Untergruppe enthält diejenigen Fälle, bei denen entweder keine Antwort erfolgte (registriert mit der Codeziffer 225 "Befragter kann Frage nicht beantworten, Informationsmangel"), oder bei denen die Antwort verweigert wurde (Codeziffer 226). Zusätzlich sind diejenigen Fälle interessant, bei denen im Fragebogen keinerlei Markierung bzw. keine Angabe (= KA) vorliegt. Mit dem verwendeten Codesystem war es möglich, solche Fälle vom Interviewverlauf her nachzuvollziehen und zu recherchieren, welche Umstände dazu führten, daß der Interviewer keine Angabe machte.

Es zeigt sich, daß der Interaktionsverlauf bei diesen Fragen zwischen Interviewer und Befragtem aus längeren Interaktionsketten besteht, die dadurch entstehen, daß der Interviewer sich zunächst bemüht, den Befragten dazu zu bewegen, eine adäquate Antwort zu geben. Bleiben diese Bemühungen jedoch erfolglos, weil der Befragte z.B. darauf beharrt, seine Antwort nur mit gewissen Einschränkungen oder unter zusätzlichen Voraussetzungen geben zu können, die im vorgegebenen Fragenstimulus nicht enthalten sind, oder sich nicht eindeutig für eine Stellungnahme entscheiden will, steht der Interviewer vor dem Dilemma, entweder gar keine Eintragung vorzunehmen oder fälschlicherweise die Kategorien "weiß nicht" bzw. "verweigert" - soweit vorhanden - zu markieren. Die Möglichkeit, handschriftliche Erläuterungen im Fragebogen vorzunehmen, ist zum einen für den Interviewer bei den Survey-Studien zu zeitaufwendig und aus ökonomischen Gründen nicht vertretbar. Sie würde sich gleichzeitig störend auf den Interviewverlauf auswirken, zum anderen wäre dies in Fragebogen, die maschinell im Belegleseverfahren ausgewertet werden, nicht zulässig. So zeigen die Erfahrungen im Umgang mit dem Codesystem, daß sich hinter dieser "keine Angabe"-Kategorie in den meisten Fällen kein eigentlicher Fehler des Interviewers verbirgt. Vielmehr handelt es sich überwiegend um das "Endprodukt" einer vorangegangenen langen Interaktion, die trotz Bemühen des Interviewers nicht zu einer adäquaten Antwort führte.

Die Ursachen für diese unerwünschten Verhaltensweisen sind jedoch nicht nur in fragenspezifischen Konstruktionsmängeln oder im Fehlverhalten des Interviewers zu sehen, sondern auch als Folge standardisierter Fragensvorgaben, die dem Befragten keine "individuelle" Reaktionsweise zugestehen, und ihn in ein vorgegebenes Antwortschema zu pressen versuchen.

ad 2. Bei der zweiten Untergruppe handelt es sich um Fälle, bei denen im Fragebogen eine Antwort verzeichnet wurde, die aber bei der Vercodung als inadäquat erfaßt wurde. Solche Fälle können einerseits inadäquate Antworten im engeren Sinne sein (Codeziffer 222 "Befragter gibt inadäquate Antwort bei geschlossener Frage"), andererseits durch Inkonsistenzen hervorgerufene invalide Antworten (wenn z.B. jemand durch Zusatzäußerungen seine Antwort in Frage stellt = Codeziffer 205).

Normalerweise werden diese Fälle bei der Auswertung gar nicht erkannt. Eine Differenzierung dieser Art vermag nur ein entsprechendes Codesystem wie das vorliegende zu leisten. Daß derartige Fälle aber durchaus in nicht unbeträchtlichem Ausmaß auftreten, zeigt Tabelle 9 mit 34.9% für Item A, 30.0% für Item E, 26.7% für Item G. Ergänzend sei erwähnt, daß auch die restlichen, hier nicht aufgeführten Items der Frage 9 einen hohen Anteil inadäquater Antworten dieser Art aufweisen. Der Grund erklärt sich fragenspezifisch und kann auf 2 Ursachen zurückgeführt werden: Die lange Itembatterie (von A - H) mit inhaltlich schwierigen Items und gleichzeitig nicht-optischer Verbalskala mit 4 Ausprägungen führte bei dieser Frage häufig dazu, daß der Befragte nicht die differenzierte Verbalisierung der Skalenpunkte benutzte, da er diese nicht in Erinnerung behalten konnte, sondern mit eigenen Worten antwortete bzw. lediglich "zustimmte". Aufgrund der durch die Vielzahl der Items langwierigen Frage unterließ der Interviewer nach anfänglich korrekter Vorgehensweise die erforderliche Rückfrage nach der genauen Skalenbezeichnung, ordnete die inadäquate Antwort selbständig zu und lieferte somit eine im Fragebogen korrekt erscheinende Markierung.

Betrachtet man bei den Items A, E und G zu Frage 9 auch den Anteil an Idealverläufen (Spalte I) sowie den Anteil von Fällen mit adäquater Antwort ohne idealen Verlauf (Spalte II), so wird deutlich, daß hier sehr ähnliche Interaktionsprozesse vorzuliegen scheinen: Ein geringer Anteil an Idealver-

laufen und relativ hohe Anteile sowohl an Fällen mit adäquater Antwort ohne idealen Verlauf als auch an Fällen mit inadäquater Antwort.

Diese Befunde lassen darauf schließen, daß Probleme bei der Beantwortung dieser Frage sich bereits in dem hohen Prozentsatz adäquater Antworten ohne idealen Verlauf niederschlagen. Darüber hinaus ist davon auszugehen, daß für alle drei Items dieser Frage ähnliche Problemstrukturen vorliegen.

Ähnliches gilt für die Demographiefrage nach der beruflichen Ausbildung in Abbildung 4.

Abbildung 4: Demographiefrage S4

INT.: Liste S2 vorlegen. Mehrfachnennungen möglich

Und nun zu Ihrer beruflichen Ausbildung: Was von dieser Liste trifft auf Sie zu?

- A bin noch in der Ausbildung
- B keinen beruflichen Ausbildungsabschluß
- C beruflich-betriebliche Anlernzeit mit Abschlußzeugnis,
aber keine Lehre
- D Lehre mit Abschlußprüfung
- E Berufliches Praktikum, Volontariat
- F Berufsschulabschluß
- G Berufsfachschulabschluß
- H Meister-, Techniker- oder gleichwertiger Fachschulabschluß
- I Fachhochschulabschluß (auch Ingenieurschulabschluß)
- K Hochschulabschluß
- L anderen beruflichen Ausbildungsabschluß,
und zwar:.....

Wie man aus Tabelle 10 ersieht, traten bei dieser Frage keine Idealverläufe auf. Dagegen wurde ein äußerst hoher Prozentsatz (91.7%) von Fällen mit adäquater Antwort ohne idealen Verlauf, d.h. mit Zusatzaktivitäten registriert sowie 8.3% Fälle mit im Fragebogen adäquat verzeichneten Antworten, die sich jedoch bei der Vercodung als invalide erwiesen. Diese Ergebnisse sind ein Indiz dafür, daß die Beantwortung dieser Frage Probleme schafft. Als oberflächlicher Beobachter könnte man argumentieren, daß in immerhin 91.7% der Fälle adäquate Antworten erfolgten und somit das für jede Frage gültige und angestrebte Ziel in ausreichend hohem Maße erfüllt wurde. Geht man jedoch, wie einführend erwähnt, davon aus, daß Eingriffe eines Inter-

viewers in den Erhebungsprozeß eine mögliche Quelle für fehlerhaftes Verhalten darstellen können und auf Nachfragen des Befragten schließen lassen, so bilden die 91.7% der Fälle mit adäquater Antwort, aber ohne idealen Verlauf ein Indiz für die mangelnde Güte dieser Frage.

Tabelle 10: Prozentsätze der Interaktionsverläufe für die Demographiefrage S4

Frage	I Ideal- verläufe	II Fälle mit adä- quater Antwort <u>ohne</u> idealen Verlauf	III Fälle mit inadäquaten Antworten		Gesamt
			missings (KA,wn,vw)*	inadäquate Antwort, im Fragebogen adäquat bezeichnet	
S4	0.0%	91.7%	0.0%	8.3%	8.3%

*KA = Keine Angabe im Fragebogen

wn = weiß nicht

vw = verweigert

Die Analyse einer "normalen" Auswertung würde keine Hinweise auf Erhebungsprobleme dieser Art bieten, da tatsächlich in allen Fällen adäquate Antworten im Fragebogen registriert wurden. Erst die Differenzierung durch das Codesystem in Interaktionsabläufe mit Zusatzaktivitäten sowie in Untergruppen für Fälle mit im Grunde inadäquater Antwort läßt die Schwierigkeiten der operationalen Definition dieser Frage erkennen. Die Gründe liegen in der mangelnden Trennschärfe der Antwortkategorien. Der Befragte findet auf der Liste nicht auf Anhieb die für ihn zutreffende Kategorie, was zu einem Dialog zwischen Interviewer und Befragtem führt, der sich in seiner Häufigkeit in dem hohen Prozentsatz in Gruppe II niederschlägt. Versuche des Interviewers, die häufig offen gegebene Antwort des Befragten selbst einzuordnen, führten in 8.3% der Fälle zu unzutreffenden Markierungen.

Die Beispiele an einzelnen Fragen zeigen, daß das Codesystem durch die differenzierte Betrachtung von Interaktionsverläufen und inadäquatem Antwortverhalten in der Lage ist, Informationen über das Funktionieren einer Frage zu liefern, die in dieser Art durch übliche Pretestmethoden nicht aufgezeigt werden können.

Im folgenden soll noch kurz auf 2 Randergebnisse eingegangen werden, die zwar nicht direkt mit der Fragestellung zusammenhängen, die jedoch interessant erscheinen.

Zur Erklärung der über den Idealverlauf hinausgehenden Zusatzaktivitäten scheint es sinnvoll zu fragen, inwieweit diese Aktivitäten von bestimmten Merkmalen des Befragten und der Interviewer abhängen. Von besonderem Interesse ist, inwieweit das unterschiedliche Ausmaß von Zusatzaktivitäten auf unterschiedliche Interviewer zurückgeht, da hiervon Hinweise zur Interviewerauswahl und Interviewerschulung erwartet werden können.

Tabelle 11: Befragtenmerkmale und durchschnittliche Anzahl von idealen Verläufen pro Interview

Befragten- merkmale	GESCHLECHT		bis 44 (N=30)	ALTER	
	männl. (N=38)	weibl.5 (N=22)		45-59 (N=20)	60 u.älter (N=10)
Anzahl der Idealverläufe	27.7	28.1	29.6	30.4	17.4

Als Befragtenmerkmale wurden lediglich die Merkmale Geschlecht und Alter betrachtet. Tabelle 11 enthält die Ergebnisse, wobei für jede Untergruppe die durchschnittliche Anzahl der idealen Verläufe über alle Fragen dargestellt ist. Zur statistischen Überprüfung der dargestellten Unterschiede wurden angesichts der niedrigen Fallzahlen nicht-parametrische Tests gewählt. Die Überprüfung der Geschlechtsunterschiede wurde mit dem U-Test von MANN-WHITNEY und dem Mediantest vorgenommen. Sie ergab, daß die Unterschiede nicht signifikant sind. Zumindest im Falle der vorliegenden Stichprobe konnten also Geschlechtsunterschiede in bezug auf die Idealverläufe nicht nachgewiesen werden.

Anders sieht es bei den Altersgruppen aus. Hier ergab die Durchführung des Tests von KRUSKAL-WALLIS eine Signifikanz auf dem 1%-Niveau. Die Signifikanz ergab sich im wesentlichen aufgrund der Unterschiede zwischen der Gruppe der über 60jährigen und den übrigen. Es scheint sich also anzudeuten, daß ältere Befragte mehr Aktivitäten hervorrufen.

Tabelle 12: Durchschnittliche Anzahl der Idealverläufe pro Interviewer und pro Interview

Interviewer-Nr.	Durchschnittl. Idealfälle pro Interviewer	Anzahl der realisierten Fälle	Standardabweichung
1	36.3	3	11.6
2	34.3	4	4.2
3	31.7	3	8.3
4	33.0	5	4.5
5	27.4	5	10.3
6	22.0	5	15.5
7	22.0	5	12.1
8	25.2	13	10.1
9	27.7	6	4.3
10	27.0	2	11.3
11	29.6	5	7.2
12	35.3	4	18.8

Tabelle 12 zeigt die Ergebnisse für die einzelnen Interviewer, wobei die Numerierung der Interviewer nicht mit der aus Untersuchung 1 identisch ist. Die Zeilen geben einerseits die durchschnittliche Anzahl der Idealverläufe pro Interviewer, andererseits die Standardabweichung der Idealverläufe wieder. Zunächst zeigt die Tabelle, daß es durchaus Unterschiede zwischen den Interviewern gibt. Einige Interviewer haben offensichtlich im Durchschnitt mehr Idealverläufe produziert als andere. Die Standardabweichungen zeigen, daß sich diese Unterschiede nicht nur auf die Mittelwerte, sondern auch auf die Streuungen beziehen. Einige Interviewer sind offensichtlich konsistenter in der Hervorbringung von Idealverläufen als andere. Betrachtet man nur diejenigen Interviewer, die etwa die gleiche Anzahl von 4-6 Interviews durchgeführt haben, so ergibt sich eine Schwankungsbreite von Streuungsunterschieden zwischen 18.8 und 4.2. Diese Ergebnisse legen den Gedanken nahe, den Einsatz eines Codesystems wie des vorliegenden zur kombinierten Bewertung von Interviewern nach durchschnittlicher Anzahl von Idealverläufen und Streuung vorzunehmen. Das Streuungsmaß erscheint mindestens ebenso wichtig wie das Durchschnittsmaß, da man sich eigentlich Interviewer wünscht, die durchgehend "sauber" arbeiten.

3.5. Bewertung des Systems zur Qualitätsbestimmung von Fragen

Zur Beantwortung der eingangs gestellten Frage läßt sich auf der Grundlage der aufgezeigten Ergebnisse folgendes sagen: Wenn wir den Qualitätsbegriff durch Interaktion und adäquate Antwort operational definieren, so zeigt sich, daß das System sehr wohl in der Lage ist, Qualitätsmerkmale aufzuzeigen, die uns etwas über das "Funktionieren" einer Frage aussagen.

Zunächst kann das System zwischen verschiedenen Interaktionsverläufen differenzieren, es kann aufzeigen, in welchem Ausmaß Zusatzaktivitäten auftreten, und es kann interessante Details bei nicht adäquatem Antwortverhalten aufdecken. Diese Qualitätsbefunde werden systematisch erarbeitet, da zur Beurteilung einer Frage die Codes bzw. Verhaltensweisen von Interviewer und Befragtem aus allen Interviews miteingehen, ein Vorteil, den die herkömmliche Pretestmethode in dieser Form nicht aufzuweisen hat. Der Interviewer berichtet dort eher unsystematisch und vereinzelt über die in seinen Interviews als relevant empfundenen Problempunkte.

Andererseits zeigt ein Vergleich der beiden Methoden, daß ganz konkrete Mängel bei einer Frage mit dem Codesystem nicht aufgedeckt werden können, da die eigentlichen Gründe für das "Nichtfunktionieren" einer Frage sich aus den numerischen Ergebnissen, d.h. den Codeziffern, nicht ablesen lassen. Dafür arbeitet hier die eingesetzte Technik zu unpräzise, es sei denn, es gilt ein Codeschema zu entwickeln, das neben allgemeinen Verhaltensbeschreibungen fragenspezifische Charakteristika berücksichtigt und damit die Erfassung solcher detaillierten Zusatzinformationen ermöglicht. Die Struktur eines solchen Codeschemas mit den darin enthaltenen Codeziffern müßte sich stark an fragebogenspezifischen Problemstellungen inhaltlicher und technischer Natur orientieren.

Insgesamt zeigen die Erfahrungen mit dem hier verwendeten Codeschema, daß die Interaction-Coding-Technik mit ihrer objektiven und systematischen Vorgehensweise und durch ihre detaillierte Betrachtung der Interaktionssequenzen einen wesentlichen Beitrag zur Bewertung der Qualität von Fragen beitragen kann, der in Kombination mit den eher qualitativen Ergebnissen konventioneller Betrachtungen zu fundierten Pretestaussagen in bezug auf die Funktionsfähigkeit eines Instruments führen kann.

Dieser Bericht zur Interaction-Coding-Technik wurde von Peter Prüfer und Margrit Rexroth verfaßt.

Literatur

- BELSON, W.A. The design and understanding of survey questions. London: Gower, 1981.
- BRENNER, M. Patterns of social structure in research interview. In: M. BRENNER (Hrsg.), Social method and social life. London, 1981, 115-158.
- BRENNER, M. Response-effects of 'role-restricted' characteristics of the interviewer. In: W. DIJKSTRA & J. von der ZOUVEN (Hrsg.), Response behaviour in the survey-interview. London: Academic Press, 1982, 131-165.
- CANNELL, Ch. & KAHN, R. Interviewing. In: G. LINDSEY & E. ARONSON (Hrsg.), Handbook of social psychology. Cambridge: Addison-Wesley, rev. ed. 1968, 526-595.
- CANNELL, Ch., LAWSON, S. & HAUSSEER, D. A technique for evaluating interviewer performance. Survey Research Center, Institute for Social Research, University of Michigan, 1977.
- CANNELL, Ch., OKSENBURG, L., CONVERSE, M. Experiments in interviewing techniques. Field experiments in health reporting, 1971-1977. Survey Research Center, Institute for Social Research, University of Michigan, 1979.
- CANNELL, Ch., MILLER, P., OKSENBURG, L. Research of interviewing techniques. In: S. LEINHARDT (Hrsg.), Sociological Methodology. San Francisco: Jossey-Bass, 1981, 389-437.
- CANNELL, Ch., KALTON, G.W., FOWLER, F.W. Techniques for diagnosing cognitive and affective problems in survey questions. Survey Research Center, Institute for Social Research, University of Michigan, 1985.
- GORDON, R. Interviewing: strategy, techniques and tactics. Homewood: The Dorsey Press, rev. ed. 1975.
- HYMAN, H. et al. Interviewing in social research. Chicago: University of Chicago Press, 1975.
- MATHIOWETZ, N., CANNELL, Ch. Coding interviewer behavior as a method of evaluating performance. Proceedings of the American Statistical Association, Survey Methods Section, 1980, 525-528.
- MORTON-WILLIAMS, J. A study of question failure through the use of interaction coding. Invited paper, 44th session of the International Statistical Institute. Madrid, 1983.
- PAYNE, S.L. The art of asking questions. Princeton, N.J.: Princeton University Press, 1951.
- SCHUMAN, H., PRESSER, S. Questions and answers in attitude surveys: Experiments on question form, wording, and context. New York: Academic Press, 1981.
- SUDMAN, S., BRADBURN, N.M. Asking questions. A practical guide to questionnaire design. San Francisco: Jossey-Bass, 1982.

Anhang 1:

Bewertung über das Codesystem zur Bewertung von Interviewerverhalten

Codegruppe 10: Korrekte Fragestellung (inclusive Vorgabe der Antwortkategorien)

<u>Code</u>	<u>Bedeutung</u>
11:	Interviewer liest Fragetext wörtlich vor oder höchstens mit geringfügigen Veränderungen, wie z.B.: 'und', 'oder', die den Bezugsrahmen nicht verfälschen.
12:	Interviewer liest Fragetext <u>geringfügig</u> verändert vor, ohne Bezugsrahmen zu verändern; es werden <u>keine</u> Schlüsselworte hinzugefügt, weggelassen oder verändert.
13:	Interviewer liest Fragetext wörtlich und <u>vollständig</u> vor, obwohl Befragter frühzeitig antwortet.

Codegruppe 20: Nicht korrekte Fragestellung

<u>Code</u>	<u>Bedeutung</u>
21:	Interviewer liest Fragetext <u>wörtlich</u> vor, die Antwortkategorien werden <u>verändert</u> .
22:	Interviewer verändert den Fragetext erheblich; es werden Schlüsselworte hinzugefügt, weggelassen, verändert.
23:	Interviewer gibt selbst die erwartete Antwort vor; Befragter hat noch nichts geantwortet.

Codegruppe 30: Klärung oder Nachfrage auf nicht direkte Weise

<u>Code</u>	<u>Bedeutung</u>
31:	Interviewer klärt mit eigenen Worten auf nicht direkte Weise die Antwort des Befragten. Dabei bedeutet 'nicht direktiv': die Nachfrage des Interviewers darf den Bezugsrahmen der Frage nicht einschränken oder ändern; auch der Bezugsrahmen der (möglichen) Antwort des Befragten sollte weder begrenzt noch verändert werden.
32:	Interviewer wiederholt den <u>gesamten Fragetext wörtlich</u> .
34:	Interviewer wiederholt oder klärt Antwort des Befragten korrekt.

ZUMA

<u>Code</u>	<u>Bedeutung</u>
35:	Interviewer klärt oder interpretiert auf Rückfrage des Befragten den Fragetext mit eigenen Worten oder mit Hilfe des Fragetextes noch einmal korrekt.
36:	Interviewer macht angemessen lange Pausen, um Gelegenheit zum Nachdenken zu geben.

Codegruppe 40: Klärung und Nachfrage auf direkte Weise

<u>Code</u>	<u>Bedeutung</u>
41:	Interviewer macht direkte Äußerungen, die nicht im Fragetext stehen. Dabei bedeutet direktiv: Interviewer gibt dem Befragten mögliche Antworten oder zusätzliche Frageninhalte vor, die nicht im Fragetext formuliert sind. Darüber hinaus sind auch Bemerkungen des Interviewers, wie z.B.: 'war's das?' als direktiv anzusehen.
42:	Interviewer wiederholt Frage oder Fragenteile nicht korrekt, d.h. er liest Frage bei nochmaligem Vorgeben nicht so vor, wie sie im Fragebogen steht.
43:	Interviewer verwendet direktiven Einleitungssatz, der nicht im Fragebogen steht.
44:	Interviewer gibt eine von der tatsächlichen Antwort des Befragten abweichende Zusammenfassung oder ordnet die Antwort eigenmächtig den vorgegebenen Antwortkategorien zu.
45:	Interviewer interpretiert/klärt Frage nicht im Sinne der ursprünglichen Fragestellung.
46:	Interviewer gibt während oder nach Vorlesen des Fragetextes <u>zusätzlichen Stimulus</u> (z.B. Erklärungen) mit <u>eigenen Worten</u> , der nicht zwangsweise direktiv, aber unnötig ist.
47:	Interviewer gibt dem Befragten keine bzw. zu kurze Gelegenheit zum Nachdenken oder Formulieren und geht im Fragebogen weiter.

ZUMA

Codegruppe 50: Sonstiges angemessenes Verhalten beim Interview

<u>Code</u>	<u>Bedeutung</u>
51:	Interviewer gibt zulässige Hilfestellung ganz allgemein zur Interviewersituation (unabhängig von einzelnen Fragen), wie z.B. Formulierungen: 'Es gibt keine richtigen und falschen Antworten.'
58:	Interviewer macht sonstige Bemerkungen bei einzelnen Fragen, die zulässig sind, z.B.: 'Soll ich die Frage noch einmal vorlesen?'

Codegruppe 60: Sonstiges unangemessenes Verhalten

<u>Code</u>	<u>Bedeutung</u>
62:	Interviewer unterbricht den Befragten
63:	Interviewer gibt persönliche Meinung oder Bewertung ab, z.B. lobt den Befragten, kritisiert Befragten, zeigt sich erstaunt, zeigt Mißfallen.
64:	Falsche technische Vorgehensweise durch den Interviewer, z.B.: läßt Befragten im Fragebogen lesen, wendet Befragungshilfen nicht nach Vorschrift an, versäumt, Befragten bei falscher technischer Anwendung zu korrigieren.
65:	Interviewer liest Interviewer-Anweisung laut vor, denkt laut.
66:	Interviewer formuliert Antwort des Befragten noch einmal grundlos mit anderen Worten.
67:	Interviewer läßt Befragten zu weit abschweifen.
68:	Interviewer benutzt dem Befragten gegenüber Fachausdrücke, wie z.B.: Rating, Items, usw.

Codegruppe 70: Verhalten, das auf dem Band nicht hörbar ist

<u>Code</u>	<u>Bedeutung</u>
71:	Interviewer schreibt die schon vorher erhaltene Antwort auf, ohne die Frage noch zu stellen.

ZUMA

<u>Code</u>	<u>Bedeutung</u>
72:	Interviewer <u>versäumt</u> , bei unzulänglicher Antwort nachzufragen <u>oder</u> den Sinn der Frage richtigzustellen.
73:	Band ist unterbrochen, Inhalte fehlen.
74:	Interviewer registriert Antwort des Befragten sinnentstellt oder unvollständig im Fragebogen.
75:	Interviewer registriert Antwort des Befragten unleserlich im Fragebogen.

Codegruppe 80: Korrekter Filtersprung

<u>Code</u>	<u>Bedeutung</u>
81:	Interviewer überspringt eine oder mehrere Fragen laut Filteranweisung.

Codegruppe 90: Filterfehler

<u>Code</u>	<u>Bedeutung</u>
91:	Interviewer liest Frage, die er hätte vorlesen müssen, nicht vor.
92:	Interviewer stellt wegzulassende Frage.

Anhang 2:

Übersicht über das Codesystem zur Bewertung der Qualität von Fragen

Komplex 1: Erster Verhaltensschritt des Interviewers vor der ersten Reaktion des Befragten

Dieser Komplex umfaßt alle Verhaltensweisen des Interviewers, die ihm als fest vorgeschriebene Aufgabenstellung vorgegeben sind, vor der ersten Reaktion des Befragten, d.h. bevor der Befragte antwortet, eine Verständnisfrage stellt oder die ihm gestellte Frage in irgendeiner Weise kommentiert, wie:

- Vorlesen der Einleitung
- Vorlesen der Frageformulierung/Antwortkategorien/Skalen/Items
- Erläuterungen zur Aufgabenstellung des Befragten (z.B. Benutzung der Listen, Anweisung zu Mehrfachnennungen u.s.w.)

I	<u>Einleitung</u>	101	Int. liest eine im Fragebogen <u>vorgegebene</u> Einleitung für eine Fragengruppe oder eine einzelne Frage <u>wörtlich</u> bzw. mit <u>geringfügigen Änderungen</u> , die den Bezugsrahmen nicht verändern, vor (geringfügige Änderungen sind z.B. Bindewörter wie "und" oder "nun")
		102	Int. formuliert selbst eine im Fragebogen <u>nicht vorgegebene nicht direktive</u> Einleitung
		103	Int. verändert eine im Fragebogen <u>vorgegebene</u> Einleitung für eine Fragengruppe oder eine einzelne Frage <u>direktiv</u> , d.h. macht erhebliche Veränderungen bei der Einleitung
		104	Int. macht eine im Fragebogen <u>nicht vorgegebene direktive</u> Einleitung, die das Ziel der Frage falsch vermittelt
		105	Int. unterläßt es, eine im Fragebogen <u>vorgegebene</u> Einleitung vorzulesen und liest ohne Einleitung bzw. Überleitung den eigentlichen Fragetext vor
II	<u>Fragetext</u>	111	Int. liest Fragetext <u>wörtlich</u> vor
		112	Int. liest Fragetext mit <u>geringfügigen Veränderungen</u> vor, ohne den Bezugsrahmen zu verändern
		113	Int. liest Fragetext mit <u>erheblichen Veränderungen</u> vor, d.h. es werden Schlüsselworte oder Teile zum Fragetext hinzugefügt, weggelassen oder verändert bzw. mehrere kleine Veränderungen angebracht, die den Frageinhalt unklar machen
III	<u>Antwortkategorien/ Skalen</u>	121	Int. liest Antwortkategorien/Skalen <u>wörtlich</u>
		122	Int. liest Antwortkategorien/Skalen mit <u>geringfügigen Veränderungen</u> vor
		123	Int. liest Antwortkategorien/Skalen mit <u>erheblichen Veränderungen</u> (analog Code 113) vor
IV	<u>Items</u>	131	Int. liest Items <u>wörtlich</u>
		132	Int. liest Items mit <u>geringfügigen Veränderungen</u>
		133	Int. liest Items mit <u>erheblichen Veränderungen</u> (analog Code 113) vor
V	<u>Anderes fehlerhaftes Verhalten beim Vorlesen des Fragetextes</u>	141	Int. liest Fragetext/Antwortkategorien/Skalen/Items <u>nicht vollständig</u> vor, da <u>Zp frühzeitig antwortet</u>
		142	Int. <u>gibt selbst erwartete Antwort</u> vor 1. Reaktion der Zp vor, indem er diese aus Anmerkungen des Befragten entnommen hat
VI	<u>Filter</u>	151	Int. stellt Frage, die er hätte überspringen müssen
		152	Int. stellt Frage nicht, die er hätte stellen müssen
<u>VII Zusatzerklärungen vor der ersten Reaktion des Befragten</u>			
		161	Int. gibt <u>adäquate Erklärungen</u> zur Aufgabe des Befragten, die zum Fragendesign gehören, nicht wörtlich im Fragebogen stehen, sondern in Form einer Intervieweranweisung (z.B. Mehrfachnennungen möglich)

- 162 Int. gibt inadäquate Erklärungen zur Aufgabe des Befragten, die zum Fragedesign gehören, nicht wörtlich im Fragebogen stehen, sondern in Form einer Intervieweranweisung (z.B. eine Nennung ist zugelassen, Int. vermittelt, daß mehrere Angaben gemacht werden können)
- 163 Int. unterläßt Erklärung zur Aufgabe des Befragten, die zum Fragedesign gehören, nicht wörtlich im Fragebogen stehen, die aber zur adäquaten Beantwortung der Frage notwendig sind (z.B. Int. unterläßt Hinweis, daß Mehrfachnennungen möglich sind)
- 164 Int. macht zusätzlich adäquate Erklärungen, die im Fragebogen nicht vermerkt sind (z.B. "sehen Sie alle Karten in Ruhe durch")
- 165 Int. macht zusätzlich inadäquate Erklärungen (z. B. Int. erweitert vor der ersten Reaktion des Befragten durch eigenmächtige Zusatzserklärungen den Fragetext)

Komplex 2: Befragtenverhalten

Komplex 2 umfaßt alle vom Befragten gezeigten Verhaltensweisen

I Angemessene Antworten

- 201 geschlossene Frage, Zp gibt angemessene Antwort
- 202 offene Frage, Zp gibt angemessene Antwort
- 203 Zp denkt laut, kommentiert die Frage
- 204 Zp korrigiert die Antwort und gibt eine andere
- 205 Zp gibt zwar adäquate Antwort, doch Ton, Zögern, Erläuterungen, Inkonsistenzen lassen auf invalide Antwort schließen
- 206 Zp kommentiert/erklärt die gegebene Antwort zusätzlich (z.B. einen Skalenwert)

II Ende der Antwort

- 207 Zp signalisiert Ende der Antwort; äußert, daß sie alles gesagt hat

III Wunsch nach Klärung bezüglich Frageinhalt und Rollenverhalten; fragenspezifisch und allgemein

- 208 Zp bittet um Klärung von Inhalt bzw. Bedeutung von Fragetext, Item, Skala oder Antwortkategorien
- 209 Zp bittet um Klärung seiner Aufgabe, fragenspezifisch (z.B. soll die Antwort offen erfolgen oder durch Antwortvorgaben)
- 210 Zp bittet um Wiederholung der gesamten Frage oder von Teilen der Frage
- 211 Zp wiederholt Frage oder Teile davon, um sich selbst den Bezugsrahmen noch einmal herzustellen
- 212 Zp bittet um Auskunft über Auftraggeber, Ziel/Zweck der Untersuchung, Auswahlverfahren, Befragtenrolle, allgemein

IV Bestätigung der Klärung durch den Interviewer

- 213 Zp bestätigt Klärung der Befragten- und Interviewerrolle, fragenspezifisch oder allgemein
- 214 Zp bejaht das Angebot des Interviewers, irgend etwas noch einmal zu erklären

V Wunsch nach Bestätigung der gegebenen Antwort

- 215 Zp fragt, ob ihre Antwort korrekt ist, versucht Rückmeldung über Antwort vom Interviewer zu erhalten

VI Bestätigung der Antwort

- 216 Zp bestätigt die bereits gegebene Antwort auf Interviewerrückfrage
- 217 Zp bestätigt eine vom Interviewer nicht korrekt wiederholte Antwort

VII Nicht abschweifendes höfliches Verhalten

- 218 Zp macht höfliche Äußerungen, die nicht abschweifen, die aber nicht direkt zur Frage gehören (z.B. "haben Sie genügend Platz?")

- VIII Abschweifungen 219 Zp gibt von sich aus Informationen, die für die Frage nicht relevant sind, versucht Diskussion, Konversation zu betreiben
 220 Zp geht auf Abschweifungen des Interviewers ein
 221 Zp widerspricht dem Klärungsversuch des Interviewers
- IX Inadäquate Antworten
 222 geschlossene Frage, Zp gibt inadäquate Antwort
 223 offene Frage, Zp gibt inadäquate Antwort
 224 Zp übt Kritik an Fragetext, Skala, Items oder Antwortkategorien
- X Nichtbeantwortung der Frage
 225 Zp kann Frage nicht beantworten (Informationsmangel, Schwierigkeiten sich zu erinnern, Antwortkategorien treffen nicht zu)
 226 Zp verweigert die Antwort
- XI Unterbrechung beim Vorlesen des Fragetextes
 227 Zp unterbricht
- XII Einschaltung Dritter
 228 andere Personen, die beim Interview anwesend sind, schalten sich in das Gespräch ein; antworten anstelle der Zielperson bzw. informieren oder korrigieren diese

Komplex 3: Interviewerverhalten nach der ersten Reaktion des Befragten

Komplex 3 umfaßt alle Verhaltensweisen des Interviewers, die nach der 1. Reaktion des Befragten stattfinden und klären:

- Rolle, Ziel und Aufgabe des Befragten allgemein und im Zusammenhang mit einzelnen Fragen
- Interviewerrolle allgemein und im Zusammenhang mit einzelnen Fragen
- Inhalt/Bedeutung einzelner Fragen

- I Fragenspezifische Klärung von Aufgabe und Inhalt
 301 Int. klärt Aufgabe/Antwort des Befragten (die nicht eindeutig zuzuordnen ist) bezüglich einer Frage korrekt
 302 Int. klärt Inhalt/Bedeutung von: Fragetext, Item, Skala, Antwortkategorien korrekt
 303 Int. wiederholt Frage oder Teile davon korrekt
- II Generelle Klärung von Rolle, Ziel, Aufgabe
 304 Int. klärt Aufgabe, Rolle, Ziel allgemein, nicht fragenspezifisch, gibt Information über Auftraggeber, Auswahlprozeduren usw.
- III Probing/Ermutigung, Antwort zu vertiefen bei offenen Fragen
 305 Int. fragt nicht direktiv mit eigenen Worten nach
 306 Int. gebraucht vorgeschriebene Probes bei offenen Fragen
 307 Int. ermutigt zu weiteren Angaben durch kurze Worte (z.B. "hm"), signalisiert Aufmerksamkeit und Ermutigung
 308 Int. gibt genügend Zeit zum Nachdenken
- IV Herstellen von Motivation/Rapport
 309 Int. macht nicht abschweifende neutrale Bemerkungen, um z.B. gute Situation zu schaffen, ist höflich

- V Kontrolle bei abschweifendem Verhalten
- 310 Int. stoppt abschweifendes Verhalten z.B. durch Vorlesen der nächsten Frage oder neutrale Bemerkungen, die den Befragten auffordern im Sinne des Interviews weiterzumachen
- VI Bestätigung einer adäquaten Antwort
- 311 Int. bestätigt die adäquate Antwort und ihre Aufzeichnung, oft nur aus Höflichkeit (z.B. "gut")
- VII Rückversicherung der gegebenen Antwort
- 312 Int. versichert sich der gegebenen Antwort durch Rückfrage
- VIII Rückfrage, ob Klärung oder Wiederholung der Frage gewünscht wird
- 313 Int. fragt, ob Zp Klärung oder Wiederholung der Frage wünscht
- IX Provozierte Abschweifungen
- 314 Int. provoziert Abschweifungen, die mit dem Frage-/Antwortprozeß nichts zu tun haben
- 315 Int. unterläßt es, Abschweifungen, die der Befragte initiiert hat, zu kontrollieren
- X Infragestellung der Glaubwürdigkeit der Umfrage durch negative Kommentare
- 316 Int. äußert sich negativ in Bezug auf Interviewerrolle, in Bezug auf Aspekte der Untersuchung oder des Forschers, macht negative Kommentare bezüglich einzelner Fragen
- XI Bewertung der Antwort des Befragten
- 317 Int. gibt persönliches Urteil in Bezug auf die Antwort des Befragten ab
- XII Direktives Verhalten, das die Antwort des Befragten beeinflusst und nicht dem Sinn der Frage entspricht
- 318 Int. interpretiert/äußert sich direktiv/unzulänglich zu Inhalt/Bedeutung von: Fragetext, Item, Skala, Antwortkategorien
- 319 Int. wiederholt Frage oder Teile davon inkorrekt
- 320 Int. gibt falsche Information über Befragtenaufgabe (fragenspezifisch und generell)
- 321 Int. gibt selbst Antwort vor
- 322 Int. formuliert Antwort des Befragten noch einmal, aber inkorrekt
- 323 Int. stellt Nachfrage, die vermuten läßt, daß Interviewer die Antwort für inadäquat oder unvollständig hält (obwohl sie korrekt ist); dazu gehört übertriebenes zu häufig gestelltes Nachfragen wie: "fällt Ihnen noch etwas ein?"
- 324 Int. geht technisch falsch vor (in Bezug auf Handhabung von Fragebogen und Material) z.B. vergißt, Kärtchen zu mischen
- XIII Unterlassen von Nachfrage/Probing
- 325 Int. versäumt nachzufragen, unterläßt Klärung der Antwort bei offenen/geschlossenen Fragen
- 326 Int. versäumt vorgeschriebene Probes zu verwenden
- 327 Int. gibt nicht genügend Zeit zum Nachdenken
- XIV Unterbrechen des Befragten
- 328 Int. unterbricht Befragten

Zusammenfassung Komplex 4: Registrierung der Antworten im Fragebogen

I. GESCHLOSSENE FRAGEN

<u>Vorgabe im Fragebogen</u>	<u>Markierung</u>
401 1 Nennung zugelassen	Int. markiert falsche Codeziffer
402 Mehrfachnennungen zugelassen	Int. markiert eine oder mehrere Antwortvorgaben <u>falsch</u> , obwohl die Anzahl der markierten Kategorien mit der Anzahl der genannten Antworten <u>übereinstimmt</u>
403 Mehrfachnennungen zugelassen	Int. markiert <u>weniger</u> Kategorien als vom Befragten angegeben
404 Mehrfachnennungen zugelassen	Int. markiert <u>mehr</u> Antwortkategorien als vom Befragten angegeben

II. OFFENE FRAGEN

405 Teile oder gesamte Mitschrift unleserlich
406 Int. verändert bei der Mitschrift die Information wesentlich, mehr, weniger oder verändert
407 Band unterbrochen
408 Band unverständlich